© 1963 by the American Psychological Association, Inc.

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 65

Aiba, T. S. Can the Absolute Threshold Be Conditioned? 233

Allen, C. K., Hill, F. A., and Wickens, D. D. The Orienting Reflex as a Function of the Interstimulus Interval of Compound Stimuli 309

Allison, J. K. See Birch, D.

Bakan, P., and Mizusawa, K. Effect of Inspection Time and Direction of Rotation on a Generalized Form of the Spiral Aftereffect 583

Bakan, P., and Weiler, E. Kinesthetic Aftereffect and Mode of Exposure to the Inspection Stimulus 319

Battig, W. F., and Brackett, H. R. Transfer from Verbal-Discrimination to Paired-Associate Learning: II. Effects of Intralist Similarity, Method, and Percentage Occurrence of Response Members 507

Bertelson, P. S-R Relationships and Reaction Times to New versus Repeated Signals in a Serial Task 478

Bertucco, R. See Peterson, L. R.

Besch, N. F., Morris, H., and Levine, S. A Comparison between Correction and Noncorrection Methods in Drive Discrimination 414

Bilodeau, E. A. See Blick, K. A.

Birch, D., Allison, J. K., and House, R. F. Extinction Performance Following Discrimination Training 148

Blick, K. A., and Bilodeau, E. A. Interpolated Activity and the Learning of a Simple Skill 515

Bourne, L. E., Jr. Long-Term Effects of Misinformative Feedback upon Concept Identification 139

Bourne, L. E., Jr., and Bunderson, C. V. Effects of Delay of Informative Feedback and Length of Postfeedback Interval on Concept Identification 1

Brackett, H. R. See Battig, W. F.

Brewer, C. L. See Peterson, L. R.

Briggs, G. E. See Naylor, J. C.

Brownstein, A. J. See Marx, M. H.

Bryden, M. P. Ear Preference in Auditory Perception 103

Bugelski, B. R., and Rickwood, J. Presentation Time, Total Time, and Mediation in Paired-Associate Learning: Self-Pacing 616

Bunderson, C. V. See Bourne, L. E., Jr.

Butter, C. M. Stimulus Generalization along One and Two Dimensions in Pigeons 339

Campbell, D. T. See Harvey, O. J.

Capaldi, E. J., Hart, D., and Stanley, L. R. Effect of Intertrial Reinforcement on the Aftereffect of Nonreinforcement and Resistance to Extinction 70

Capaldi, E. J., and Stanley, L. R. Temporal Properties of Reinforcement Aftereffects 169

Capaldi, E. J., and Wargo, P. Effect of Transitions from Nonreinforced to Reinforced Trials under Spaced-Trial Conditions 318

Chang, J.-J. See Shepard, R. N.

Clark, B., and Graybiel, A. Perception of the Postural Vertical in Normals and Subjects with Labyrinthine Defects 490

Cohen, B. H. An Investigation of Recoding in Free Recall 368

Corah, N. L., and Stern, J. A. Stability and Adaptation of Some Measures of Electrodermal Activity in Children 80

Cornwell, H. G. Prior Experience as a Determinant of Figure-Ground Organization 156

Cotton, J. W., and Jensen, G. D. Successive Acquisitions and Extinctions in a T Maze 546

Cross, H. A. See Wickens, D. D.

Deane, G. E. See Jenks, R. S.

Denny, J. P. See Spielberger, C. D.

Derks, P. L. Effect of Run Length on the "Gambler's Fallacy" 213

Duncan, C. P. Effect of Instructions and Information on Problem Solving 321

Ebel, H. C., and Prokasy, W. F. Classical Eyelid Conditioning as a Function of Sustained and Shifted Interstimulus Intervals 52

Ebenholtz, S. M. Position Mediated Transfer between Serial Learning and a Spatial Discrimination Task 603

Eisler, H., and Ottander, C. On the Problem of Hysteresis in Psychophysics 530

Feldman, S. E. Probabilistic Hierarchies to Ambiguous Concept Classes 240

Fitts, P. M., Peterson, J. R., and Wolpe, G. Cognitive Aspects of Information Processing: II. Adjustments to Stimulus Redundancy 423

Fitzgerald, M. L. See Martin, J. G.

Fleshler, M. See Hoffman, H. S.

Foote, W. E. See Havens, L. L.

Fowler, H. See Wischner, G. J.

FRIEDMAN, H. Wave-Length Generalization as a Function of Spacing of Test Stimuli....... 334

GANZ, L. Effect of an Anchor Stimulus on the Stimulus Generalization Gradient........... 270

GEER, J. See LANG, P. J.

GOGGIN, J. Influence of the Written Recall Measure on First-List Associations............. 619

GOLDSTEIN, H., AND SPENCE, K. W. Performance in Differential Conditioning as a Function of Variation in Magnitude of Reward.. 86

GORMEZANO, I. See MOORE, J. W.

GRAYBIEL, A. See CLARK, B.

GREGG, L. W., AND KARN, H. W. Perceptual Responses as a Function of the Sequential Properties of Multiple Visual Stimuli.. 124

GRICE, C. R., SIMMONS, H. J., AND HUNTER, J. J. Failures to Obtain Mediated Generalization Effects in Eyelid Conditioning.. 485

HABER, A. See KALISH, H. I.

HABER, R. N. See HARRIS, C. S.

HARRIS, C. S., AND HABER, R. N. Selective Attention and Coding in Visual Perception...... 328

HART, D. See CAPALDI, E. J.

HARTMAN, T. F. Semantic Transfer of the Differential Conditioned Eyelid Response from Words to Objects.. 194

HARVEY, O. J., AND CAMPBELL, D. T. Judgments of Weight as Affected by Adaptation Range, Adaptation Duration, Magnitude of Unlabeled Anchor, and Judgmental Language........ 12

HARWAY, N. I. Judgment of Distance in Children and Adults.. 385

HAVENS, L. L., AND FOOTE, W. E. The Effect of Competition on Visual Duration Threshold and Its Independence of Stimulus Frequency.. 6

HELLYER, S. Stimulus-Response Coding and Amount of Information as Determinants of Reaction Time.. 521

HETHERINGTON, M., AND ROSS, L. E. Effect of Sex of Subject, Sex of Experimenter, and Reinforcement Condition on Serial Verbal Learning.. 572

HIGGINS, J., MEDNICK, S. A., AND TAYLOR, S. L. A Replication of Facilitation of Concept Formation through Mediated Generalization.. 421

HILL, F. A. See ALLEN, C. K.

HILL, W. F., AND SPEAR, N. E. Extinction in a Runway as a Function of Acquisition Level and Reinforcement Percentage.. 495

HILL, W. F., AND SPEAR, N. E. A Replication of Overlearning and Reversal in a T Maze.. 317

HINKLE, A. E. See MARTIN, J. G.

HISS, R. H. See THOMAS, D. R.

HISS, R. H., AND THOMAS, D. R. Stimulus Generalization as a Function of Testing Procedure and Response Measure.. 587

HNATIOW, M. See LANG, P. J.

HOFFMAN, H. S., AND FLESHLER, M. Discrimination and Stimulus Generalization of Approach, of Avoidance, and of Approach and Avoidance during Conflict.. 280

HOROWITZ, L. M., AND IZAWA, C. Comparison of Serial and Paired-Associate Learning...... 352

HOROWITZ, L. M., AND LARSEN, S. R. Response Interference in Paired-Associate Learning.. 225

HOUSE, R. F. See BIRCH, D.

HOWELL, W. C., AND KREIDLER, D. L. Information Processing under Contradictory Instructional Sets.. 39

HUNTER, J. J. See GRICE, C. R.

IZAWA, C. See HOROWITZ, L. M.

JENKS, R. S., AND DEANE, G. E. Human Heart Rate Responses during Experimentally Induced Anxiety: A Follow-up.. 109

JENSEN, G. D. Preference for Bar Pressing over "Freeloading" as a Function of Number of Rewarded Presses.. 451

JENSEN, G. D. See COTTON, J. W.

JOHNSTON, W. A. See LLOYD, K. E.

JUNG, J. Effects of Response Meaningfulness (m) on Transfer of Training under Two Different Paradigms.. 377

KALISH, H. I., AND HABER, A. Generalization: I. Generalization Gradients from Single and Multiple Stimulus Points. II. Generalization of Inhibition.. 176

KANFER, F. H. See MARSTON, A. R.

KANFER, F. H., AND MARSTON, A. R. Human Reinforcement: Vicarious and Direct......... 292

KANOTI, G. A. See KAUSLER, D. A.

KAPLAN, S. See KLEINSMITH, L. J.

KARLIN, L., AND MORTIMER, R. G. Effect of Verbal, Visual, and Auditory Augmenting Cues on Learning a Complex Motor Skill.. 75

KARN, H. W. See GREGG, L. W.

KASWAN, J., AND YOUNG, S. Stimulus Exposure Time, Brightness, and Spatial Factors as Determinants of Visual Perception.. 113

Kausler, D. A., and Kanoti, G. A. R-S Learning and Negative Transfer Effects with a Mixed List 201
Kehoe, J. Effects of Prior and Interpolated Learning on Retention in Pigeons 537
Kimmel, E., and Kimmel, H. D. A Replication of Operant Conditioning of the GSR 212
Kimmel, H. D. See Kimmel, E.
Kimmel, H. D., and Pennypacker, H. S. Differential GSR Conditioning as a Function of the CS-UCS Interval 559
King, D. J., and Tanenbaum, S. Comparison of Two Procedures in the Study of Retroactive Interference in Connected Meaningful Material 420
Kleinsmith, L. J., and Kaplan, S. Paired-Associate Learning as a Function of Arousal and Interpolated Interval 190
Klemmer, E. T. Perception of Linear Dot Patterns 468
Kothurkar, V. K. Effect of Stimulus Response Meaningfulness on Paired-Associate Learning and Retention 305
Kreidler, D. L. See Howell, W. C.
Kushnick, S. A. See Wischner, G. J.
Lachman, R., and Sanders, J. A. Concept Shifts and Verbal Behavior 22
Lang, P. J., Geer, J., and Hnatiow, M. Semantic Generalization of Conditioned Autonomic Responses 552
Larsen, S. R. See Horowitz, L. M.
Levine, S. See Besch, N. F.
Lindley, R. H. Association Value, Familiarity, and Pronunciability Ratings as Predictors of Serial Verbal Learning 347
Lloyd, K. E., and Johnston, W. A. Short-Term Retention as a Function of Contextual Constraint 460
Loewe, E. M. See Martin, J. G.
London, P. See Rosenhan, D.
McAllister, D. E. See McAllister, W. R.
McAllister, W. R., and McAllister, D. E. Increase over Time in the Stimulus Generalization of Acquired Fear 576
McBurney, D. H., and Pfaffmann, C. Gustatory Adaptation to Saliva and Sodium Chloride . . 523
McCullers, J. C. An Analysis of Some Factors Underlying Intralist Associative Transfer in Paired-Associate Learning 163
McHose, J. H. Effect of Continued Nonreinforcement on the Frustration Effect 444
Marston, A. R. See Kanfer, F. H.
Marston, A. R., and Kanfer, F. H. Group Size and Number of Vicarious Reinforcements in Verbal Learning 593
Martin, C. J., and Saltz, E. Serial versus Random Presentation of Paired Associates 609
Martin, J. G., Loewe, E. M., Hinkle, A. E., and Fitzgerald, M. L. Net Approach Gradient in Approach-Avoidance Conflict 520
Marx, M. H. See Pieper, W. A.
Marx, M. H., and Brownstein, A. J. Effects of Incentive Magnitude on Running Speeds without Competing Responses in Acquisition and Extinction 182
Meade, R. D. Effect of Motivation and Progress on the Estimation of Longer Time Intervals . . 564
Mednick, S. A. See Higgins, J.
Mizusawa, K. See Bakan, P.
Moore, J. W., and Gormezano, I. Effects of Omitted versus Delayed UCS on Classical Eyelid Conditioning under Partial Reinforcement 248
Morris, H. See Besch, N. F.
Morrissette, J. O., and Pearson, W. H. Prediction of Behavior under Conditions of Uncertainty 391
Mortimer, R. G. See Karlin, L.
Murdock, B. B., Jr. Short-Term Retention of Single Paired Associates 433
Myers, J. L., and Myers, N. A. Secondary Reinforcement in Children as a Function of Conditioning Associations, Extinction Percentages, and Stimulus Types 455
Myers, N. A. See Myers, J. L.
Naylor, J. C., and Briggs, G. E. Effects of Task Complexity and Task Organization on the Relative Efficiency of Part and Whole Training Methods 217
Ottander, C. See Eisler, H.
Pearson, W. H. See Morrissette, J. O.
Pennypacker, H. S. See Kimmel, H. D.
Peterson, J. R. See Fitts, P. M.
Peterson, L. R., Brewer, C. L., and Bertucco, R. A Guessing Strategy with the Anticipation Technique 258
Pfaffmann, C. See McBurney, D. H.
Pieper, W. A., and Marx, M. H. Effects of Within-Session Incentive Contrast on Instrumental Acquisition and Performance 568

Poulton, E. C., and Simmonds, D. C. V. Value of Standard and Very First Variable in Judgments of Reflectance of Grays with Various Ranges of Available Numbers 297
Prokasy, W. F. See Ebel, H. C.
Reynolds, J. H. Repetition Factors in One-Trial Learning of Paired Associates 352
Rickwood, J. See Bugelski, B. R.
Rosenhan, D., and London, P. Hypnosis in the Unhypnotizable: A Study in Rote Learning .. 30
Ross, L. E. See Hetherington, M.
Rotberg, I. C., and Woolman, M. Verbal Paired-Associate Learning as a Function of Grouping Similar Stimuli or Responses 47
Runquist, W. N. Performance in Eyelid Conditioning Following Changes in Reinforcement Schedule 617
Saltz, E. See Martin, C. J.
Sanders, J. A. See Lachman, R.
Shaffer, J. P. Effect of Different Stimulus Frequencies on Discrimination Learning with Probabilistic Reinforcement 265
Shepard, R. N., and Chang, J.-J. Stimulus Generalization in the Learning of Classifications .. 94
Simmonds, D. C. V. See Poulton, E. C.
Simmons, H. J. See Grice, C. R.
Smith, O. W., and Smith, P. C. On Motion Parallax and Perceived Depth 107
Smith, P. C. See Smith, O. W.
Solley, C. M. See Thomas, A.
Spear, N. E. See Hill, W. F.
Spence, K. W. See Goldstein, H.
Spence, K. W., and Tandler, B. F. Differential Eyelid Conditioning under Equated Drive as a Function of the Reinforcing UCS 35
Spielberger, C. D., and Denny, J. P. Visual Recognition Thresholds as a Function of Verbal Ability and Word Frequency 597
Stanley, L. R. See Capaldi, E. J.
Stern, J. A. See Corah, N. L.
Tandler, B. F. See Spence, K. W.
Tanenbaum, S. See King, D. J.
Taylor, S. L. See Higgins, J.
Teichner, W. H. Perception and Short-Term Memory in the Identification of Multisymbol Displays 407
Thomas, A., and Solley, C. M. Search-Discrimination Time for Missing Stimulus Information 501
Thomas, D. R. See Hiss, R. H.
Thomas, D. R., and Hiss, R. H. A Test of the "Units Hypothesis" Employing Wave-Length Generalization in Human Subjects 59
Trabasso, T. R. Stimulus Emphasis and All-or-None Learning in Concept Identification 398
Wagner, A. R. Sodium Amytal and Partially Reinforced Runway Performance 474
Wargo, P. See Capaldi, E. J.
Waugh, N. C. Two Methods for Testing Serial Memorization 215
Weiler, E. See Bakan, P.
Wells, H. Effects of Transfer and Problem Structure in Disjunctive Concept Formation 63
Wickens, D. D. See Allen, C. K.
Wickens, D. D., and Cross, H. A. Resistance to Extinction as a Function of Temporal Relations during Sensory Preconditioning 206
Wischner, G. J., Fowler, H., and Kushnick, S. A. Effect of Strength of Punishment for "Correct" or "Incorrect" Responses on Visual Discrimination Performance 131
Wolpe, G. See Fitts, P. M.
Woolman, M. See Rotberg, I. C.
Young, S. See Kaswan, J.
Zipf, S. G. Effects of Probability of Reward and Speed Requirement on Human Performance. 106

Journal of
Experimental Psychology

Vol. 65, No. 1 January 1963

EFFECTS OF DELAY OF INFORMATIVE FEEDBACK AND LENGTH OF POSTFEEDBACK INTERVAL ON CONCEPT IDENTIFICATION[1]

LYLE E. BOURNE, Jr. and C. VICTOR BUNDERSON

University of Utah

The purpose of this experiment was to determine the independent effects of delay of informative feedback (IF), length of post-IF interval, and task complexity on concept identification. The 216 Ss served individually in a factorial design which combined 3 delays, 3 post-IF intervals, and 2 levels of task complexity. All Ss were given standard oral instructions about the task and served to a criterion performance of 32 consecutively correct stimulus identifications. Results indicated that task difficulty increased with complexity, performance improved linearly with increases in post-IF interval, increase in the post-IF interval was more facilitative in tasks of greater complexity, and delay of IF was an ineffective variable. The data require detailed modifications of a mathematical theory of concept identification.

The results of a recent concept identification study (Bourne, 1957) strongly suggested the existence of an inverse relationship between performance and delay of informative feedback (IF). These findings appeared to be consistent with most previous data on the effects of delay of IF (e.g., Greenspoon & Foreman, 1956; Saltzman, 1951) and delay of reinforcement (e.g., Grice, 1948; Spence, 1947) in a variety of tasks, and a general theoretical analysis, based on the decay or forgetting of relevant stimulus cues during delay, was proposed (Bourne, 1957; Bourne & Restle, 1959). Several studies which have since been reported (Bilodeau & Bilodeau, 1958; Bilodeau & Ryan, 1960; Denny, Allard, Hall, & Rokeach, 1960; Noble & Alcock, 1958), however, failed to find any deleterious effect of delayed IF on human learning and provided data which seriously question the validity of any generalization from the performance of animals with delayed reinforcement to human behavior under conditions of delayed IF.

In four of five studies of simple motor learning, Bilodeau and Bilodeau found length of unfilled delay to be an ineffective variable; in the fifth experiment, delayed IF *facilitated* performance. The critical factor in their experiments was intertrial interval, performance varying inversely with its length. The study of Denny et al., which used a different motor skills task, provided additional support for this conclusion. Bilodeau and Bilodeau further reported that the interval separating IF from the presentation of the next stimulus, i.e., the post-feedback (post-IF) interval, might *retard* performance if it were sufficiently long.

---

[1] This research was supported in part by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service.

The apparent discrepancy between the concept identification findings and the more recent results may be attributable to a number of methodological factors. Studies in which delayed IF has been ineffective used simple motor tasks such as line drawing or knob turning primarily. Such tasks differ from concept identification in a number of dimensions, one of which being task difficulty, and Bourne's results show that task difficulty is potentially an important factor determining the influence of delayed IF. Another possibility, however, is that Bourne's results were not due to IF delay but to the post-IF interval, which was confounded with it in his experiment. In order to keep a constant inter-stimulus interval in all conditions, post-IF duration was decreased (9 to 1 sec.) to compensate for increases in delay (0 to 8 sec.). If the suggestion of Bilodeau and Bilodeau, that delay of IF is less important than the post-IF interval, is true for concept identification, then Bourne's results were less a function of the former than of the latter variable. Note, however, that if this is the case, these results are contrary to the hypothesis that increases in post-IF interval retard performance.

The purpose of this experiment was to determine the independent effects of IF delay and post-IF intervals on performance in concept identification tasks. The experimental design varied these factors orthogonally within the range used in Bourne's study. Included within the design were six experimental conditions of delay and post-IF lengths from the previous study, three at each of two levels of task complexity. These conditions were replicated to permit close comparison with earlier results.

It should be noted that, if the results of the present experiment fail to support the assumed importance of delayed IF, they will necessitate a major revision in the theory of concept identification proposed by Bourne and Restle, at least in so far as that theory interprets the effect of delayed IF. The theory, as it stands, makes no prediction about the effect of variation in post-IF interval. For a detailed account of predictions regarding task complexity and delay of IF, see Bourne and Restle (1959).

PROCEDURE

Subjects.—The *S*s, 216 students from elementary psychology courses, were awarded class credit for their participation. Each was assigned to 1 of the 18 treatments of the experiment and served individually for one session. All *S*s were given detailed oral instructions about the task, the meaning of IF signals, and the criterion of problem solution.

Task and apparatus.—The task and apparatus were essentially those described in detail by Bourne (1957). Briefly, *S* was presented with a series of geometric designs, each of which was a combination of the levels of two relevant and either one or five irrelevant binary stimulus dimensions such as form (square-triangle) and orientation (upright-tilted figures). To identify the category to which each design belonged, *S* pressed one of four unlabeled response buttons. The *S*'s task was to discover that these buttons corresponded to the four combinations of the levels of dimensions relevant to solution. After each response, *S* was informed of the correct category for that stimulus by the onset of an IF light over the appropriate button. As in earlier studies, irrelevant dimensions appeared at each of their two levels within the sequence of stimulus designs but were uncorrelated with IF signals. The *S* responded to each stimulus after a self-determined interval and performed until he reached the criterion of problem solution, 32 consecutively correct responses.

In addition to the response and IF panel, there were four major units of apparatus: (*a*) a Dunning animatic 16-mm. filmstrip projector which presented the stimulus designs; (*b*) an Esterline-Angus operations recorder which recorded *S*'s response and the correct category for each stimulus; (*c*) a timing unit which controlled the delay and the duration (1 sec.) of IF and the length of the post-IF interval; and (*d*) a Western Union tape transmitter equipped with a precoded tape which determined the sequence of IF signals. On any trial, *S*'s response to the given stimulus advanced the projector to a blank frame, activated the timing unit, and was recorded on the Esterline-Angus. After a delay of 0 to 8 sec., the timing unit advanced the tape transmitter which presented the appropriate IF and, after 1 sec., advanced the transmitter again to turn off the signal light. After an interval of 1 to 9 sec., the timing unit advanced the projector to the next stimulus frame.

Design.—A 3 × 3 × 2 factorial design was used with three delays of IF (0, 4, and 8 sec.), three lengths of post-IF interval (1, 5, and 9 sec.), and two degrees of task complexity (1 and 5 irrelevant stimulus dimensions). The three combinations of delay and post-IF interval in this design which appeared also in Bourne's (1957) experiment are 0-9, 4-5, and 8-1, where the first number refers to delay and the second to post-IF in seconds, respectively. In addition, three problems, differing in relevant dimensions, were used in each of the 18 treatment conditions to offset the possibility of collusion among *S*s. This variable was orthogonal to the others so that its effect on performance could be evaluated.

RESULTS AND DISCUSSION

An analysis of variance, summarized in Table 1, was performed on total errors to solution. Three main effects, task complexity (number of irrelevant dimensions), problems, and length of post-IF interval, were statistically significant. All three two-way interactions and the three-way interaction of these variables were also significant. A similar analysis, with trials to solution as the measure, yielded the same results except that the Problems × Complexity interaction was not reliable.

TABLE 1

ANALYSIS OF VARIANCE ON NUMBER OF ERRORS

Source	*df*	*F*
Postfeedback interval (PIF)	2	11.83**
Linear	1	23.64**
Quadratic	1	.02
Delay of IF (DIF)	2	1.77
Problems (P)	2	12.82**
Task complexity (C)	1	103.47**
PIF × DIF	4	.26
PIF × P	4	4.70**
PIF × C	2	6.21**
DIF × P	4	.35
DIF × C	2	1.67
P × C	2	4.88**
PIF × DIF × P	8	1.16
PIF × DIF × C	4	1.29
PIF × P × C	4	3.52**
DIF × P × C	4	.33
PIF × DIF × P × C	8	1.21
Residual *MS*	162	(677.22)

** $p < .01$.

The effect of task complexity is consistent with most preceding concept identification experiments (e.g., Bourne, 1957), in which the relationship of this variable to performance has been linear. Mean numbers of errors for one and five irrelevant dimensions were 14.57 and 50.33, respectively, in the present experiment. The effect of problem differences, as in earlier studies (e.g., Brown & Archer, 1956), is attributable to the difficulty *S*s have when positional dimensions, such as horizontal location of the design on the viewing screen, are relevant to solution.

The unique and more important results of the present experiment concern the effects of IF delay and length of post-IF interval. Neither delay of IF nor any of its interactions with other variables was a significant source of variance in the analyses. Increase in post-IF interval, on the other hand, produced a uniform linear decrease in mean number of errors, as shown by an orthogonal polynomial analysis. The observed effect of variation in post-IF interval was greater in problems with five than in problems with one irrelevant dimension. Figure 1 displays graphically the effects of delay, post-IF interval, and task complexity. Note the data points from conditions common to this and to Bourne's study of delayed IF. Figure 1 clearly indicates that the apparent increase in mean number of errors with IF delay in the earlier study was an artifact of the confounding of delay with post-IF interval. Actually, performance improved slightly but not significantly in the present experiment with independent increases in IF delay, especially in problems with five irrelevant dimensions.

The outcome of the present experiment is similar to findings in studies of simple motor learning and supports the conclusion that a short unfilled delay interval between response and IF exerts little or no effect on the performance of human *S*s. The argument that the difference in findings reported by Bourne (1957) and by Bilodeau and Bilodeau (1958) might be attributable to methodological or task complexity factors was not supported, although these variables are probably critical in determining the exact nature of effects of temporal characteristics. The present findings are at variance with the conclusions of Bilodeau and Bilodeau in that performance improved linearly, rather than worsened, as the post-IF interval increased. Greater lengths of post-IF interval *may* produce the opposite effect, but it will be necessary to undertake additional experiments to determine whether an optimal post-IF interval exists.

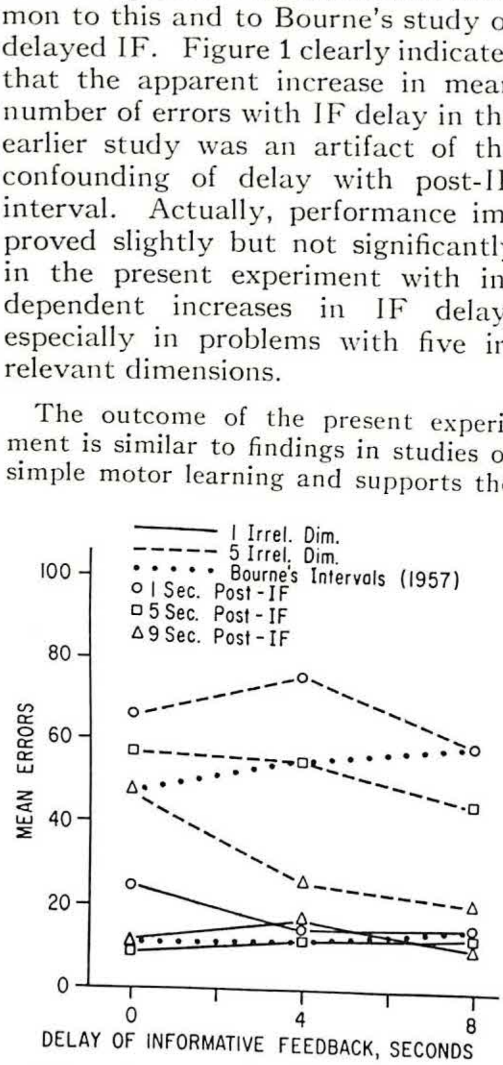


FIG. 1. Mean number of errors to solution as a joint function of delay of IF, length of post-IF interval, and task complexity. (Each plotted point represents the data from 12 *S*s. Note the experimental conditions common to this and to Bourne's, 1957, study.)

If it were proposed that learning takes place immediately and automatically at the time of IF (reinforcement), there appears to be no reason why length of post-IF interval, as distinct from any other time interval, should affect performance. As alternatives, it might be argued that much of *S*'s behavior during the post-IF interval can be described as (*a*) processing (or simply rehearsing) information retained from the stimulus and the IF signal, or (*b*) conditioning (and adapting) stimulus elements. Then it may be hypothesized that, because *S* can process only a fraction of the available information in any short interval (Miller, 1956) or because stimulus cues fluctuate in and out of the sample effective at any instant (Estes, 1955), intratrial learning and, therefore, overall performance is more efficient with longer post-IF intervals. Investigations of the distribution of *S*-determined post-IF intervals (i.e., where *S* controls the onset of the next stimulus) may provide some useful evidence regarding these possibilities. If it is assumed that the onset of the next stimulus terminates learning on the preceding trial, it would be predicted that (*a*) as amount of stimulus information (from relevant or irrelevant dimensions) increases, *S*-determined post-IF intervals will increase and (*b*) there will be a gradual decrease in length of post-IF interval with practice since each succeeding trial reduces the amount of information or elements to be learned. Note that the first of these predictions is consistent with the Post-IF × Task Complexity interaction obtained in the present experiment.

The results of this experiment do not support the delayed IF hypothesis of Bourne and Restle (1959) and necessitate some modification in their model. The available data do not permit the conclusion that IF delay is generally ineffective; considerably longer or even short *filled* delays may produce the expected result. However, the model obviously predicts a stronger effect than is compatible with these data. To provide a more precise account of the present evidence, a parameter which is an increasing function of length of post-IF interval and a decreasing function of amounts of relevant and irrelevant stimulus information must be considered. No attempt is made to develop appropriate modifications of the theory here. Actually, a number of forms of this parameter would fit the obtained results and there is little basis for a selection among them. Additional experiments will be necessary to determine the correctness of alternative descriptions.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

BILODEAU, E. A., & RYAN, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, **59**, 414–419.

BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

DENNY, M. R., ALLARD, M., HALL, E., & ROKEACH, M. Supplementary report: Delay of knowledge of results, knowledge of task, and intertrial interval. *J. exp. Psychol.*, 1960, **60**, 327.

ESTES, W. K. Statistical theory of distributional phenomena in learning. *Psychol. Rev.*, 1955, **62**, 369–377.

GREENSPOON, J., & FOREMAN, S. Effect of delay of knowledge of results on learning a motor task. *J. exp. Psychol.*, 1956, **51**, 226–228.

GRICE, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

NOBLE, C. E., & ALCOCK, W. T. Human delayed-reward learning with different lengths of task. *J. exp. Psychol.*, 1958, **56**, 407–412.

SALTZMAN, I. J. Delay of reward and human verbal learning. *J. exp. Psychol.*, 1951, **41**, 437–439.

SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.

(Received February 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 6–11

THE EFFECT OF COMPETITION ON VISUAL DURATION THRESHOLD AND ITS INDEPENDENCE OF STIMULUS FREQUENCY[1]

LESTON L. HAVENS AND WARREN E. FOOTE

Massachusetts Mental Health Center, Boston Psychopathic Hospital

The effect of prior word usage on visual duration thresholds has been studied under tachistoscopic conditions permitting control of competing or interfering responses. The degree of competition among these responses was significantly correlated with threshold. Stimulus frequency was not a significant variable under these conditions. We conclude that visual duration thresholds are not primarily a function of the frequency of prior usage of stimuli but of the ability or inability of the stimuli to evoke high frequency competitive responses.

Over the past 15 yr. experimental studies of perception have increasingly explored the organism's contribution to perceptual achievement. A host of experiments have investigated the place of motives, attitudes, and interests in perception and have resulted in such concepts as perceptual defense (McGinnies, 1949) and vigilance. With growing methodological sophistication the importance of the *S*'s prior experience with the stimuli was demonstrated, and in particular the role of word frequency and its effect upon recognition thresholds. But just as the frequency findings explained at least part of the threshold differences attributed to perceptual defense, so there arose a need to explain the effect of frequency.

Solomon and Postman (1952) have proposed an explanation and Bruner (1957) has presented a general theory of threshold differences. Both emphasize the importance of competition between responses as a determinant of threshold.

Solomon and Postman (1952) write:

> When a stimulus pattern is presented at short durations or at low illumination intensities, only fragments of the total word stimulus are "effective." Such a stimulus fragment may be considered to represent a point on the generalization dimension of stimulus patterns capable of eliciting the correct verbal response. A given stimulus fragment may, of course, be located on several generalization dimensions, each involving a different word. Which verbal response will be given depends on the relative strengths of association which have been established, through generalization, between the particular stimulus fragment and the different response words. If the visually presented stimulus word has had a greater frequency of prior usage than any of the competing response words, a correct response is highly probable.
>
> Words of lower prior exercise frequency will be interfered with by words of higher exercise frequency. This interference will manifest itself in the tendency for *S*'s "guesses" to be high frequency words. If the actual stimulus word is a low frequency word, effective stimulus fragments will elicit erroneous "guesses" until the amount of effective stimulation becomes great enough on successive exposures to reduce the number of competing word responses (pp. 199–200).

The strength of association between any fragment and response is determined by two factors implicit in Solomon and Postman's (1952) explanation which can be usefully separated here. The first of these is a function of where on the "generaliza-

---

[1] This work was sponsored by the Foundations' Fund for Research in Psychiatry. The authors wish to thank John E. Alman for his statistical assistance, and Dorothea J. Crook for her suggestions.

tion dimension of stimulus patterns capable of eliciting the correct verbal response" the fragment falls. There will be a greater strength of association between fragment (stimulus) and response if the fragment lies close to the stimulus pattern most evocative of the response. In addition, however, strength of association depends upon the strength of the relationship between the most evocative stimulus pattern and the response. It is the strength of this relationship which Solomon and Postman suggest is at least partly a function of prior experience.

The concept of generalization dimension not only permits the gradation of stimulus fragments in order of their closeness to the most evocative stimulus pattern; it also explains why a fragment may be capable of eliciting more than one response. The fragment may be a common element on several generalization dimensions each of which is associated with a different response. Just as many stimuli can elicit the same response if each is at some point on the generalization dimension for that response, so one stimulus can elicit many responses if the responses have that stimulus point in common. It is this "buckshot" effect of tachistoscopic stimulation that opens the way for competition between responses. The fragment fires into many generalization dimensions, stimulating a variety of possible responses. Ultimately, therefore, the likelihood of any given response depends upon its relative competitive strength.

Bruner's (1957) conception of the mechanism underlying perceptual failure is in terms of the "accessibility of categories," which corresponds to association strength in the explanation of Solomon and Postman. Categories in turn correspond to generalization dimensions. In discussing the causes of interference or competition, Solomon and Postman emphasize those to do with stimulus fragmentation and frequency of usage, while Bruner suggests that the state of readiness of the organism makes interference possible through a broadening of category acceptance limits. Both explanations agree in proposing that threshold differences are a function of some competitive or interference process.

It is toward an understanding of this competitive or interference process that the present investigation is directed. We have undertaken first to demonstrate its existence; previously it has been largely inferred from response time data and not directly demonstrated. Second, we have undertaken to measure its effect on threshold and, finally, to determine the relationship between competition and word frequency.

METHOD

Stimulus materials.—Sixteen four-letter words were chosen for presentation in a Gerbrands mirror tachistoscope. The stimuli were selected on the basis of principles derived from an earlier study, to be reported elsewhere. Briefly, however, our problem was the control of competition, that is, repetitive occurrences of incorrect prerecognition responses. We had observed that stimuli which elicited a large number of incorrect prerecognition responses were, first, words capable of being made into other words with few letter changes and, in addition, words which elicited high frequency responses. Other evidence that such responses tend to be of relatively high frequency is already in the literature (McGinnies, Comer, & Lacey, 1952; Newbigging, 1961).

More specifically, our earlier work suggested that any response which, in addition to being of high frequency, remained faithful to the first and last letters of the stimulus, differing from it only with respect to a form similar middle letter, would be a strong competitor. For example, the word "list" will become a preoccupation to anyone attempting to see the word "lint" under tachistoscopic conditions, the ready substitu-

TABLE 1

BREAKDOWN OF STIMULI ACCORDING TO THE MAIN VARIABLES

Main Variables			
High Compet. (C_1) High Freq. (F_1)	High Compet. (C_1) Low Freq. (F_0)	Low Compet. (C_0) High Freq. (F_1)	Low Compet. (C_0) Low Freq. (F_0)
Stimulus Words			
desk A beat AA some AA case AA	lint 2 tusk 2 scow 3 mare 14	over AA wear AA kind AA book AA	wren 10 nave 2 fief 1 drab 4

Note.—Frequency values from the G count are: A at least 50 per million but not so many as 100 per million. AA 100 or over per million.

Numerals indicate counted occurrence per million.

tion of the s for n generating the competitor. Words such as "line" or "lift," both high frequency responses, are less probable because they do not match the form of the stimulus closely enough.

Stimuli were thus selected to control the range of prerecognition responses (clearly there are many other ways); the purpose was to determine the relationship between competitive responses and recognition threshold. Eight of the stimuli were predicted to elicit strong competitive responses (high frequency words which closely matched the stimulus) and were designated C_1. An equal number of other words were selected to which there would be few if any competitive responses; these were designated C_0. Because of the interest in the effect of stimulus frequency on threshold and the relation between competition and frequency, half of each group of eight words was of low and the other half of high frequency as defined by the Thorndike-Lorge G Count. An A and AA rating was taken as evidence of high frequency while a rating of 14 or less was used as evidence of low frequency. The letters were printed in lower case type by an IBM electric typewriter on bond paper. Eight of the words had high letters at beginning and end while the remaining eight were all even in height of letters. Table 1 illustrates the breakdown of the stimuli into groups according to the main variables.

Procedure.—The stimuli were presented in accordance with a balanced Latin square design which permitted each stimulus to proceed and follow every other stimulus once.[2] Each word was exposed three times at 10 msec., three times at 20 msec. and so on until the correct responses occurred twice successively within one time period. All responses were recorded by *E* after each exposure and the duration of the exposures was also noted. The experiment was carried out in two separate 1-hr. sessions scheduled approximately 1 wk. apart.

Subjects.—Eight *S*s were obtained from nearby college placement services and paid for their participation. Upon arrival at the hospital they were taken to the experimental room and seated in front of the tachistoscope. They were then read the following instructions:

> This is a study in visual recognition, that is, how you set about recognizing things. The machine before you will expose meaningful words for varying intervals, at first too rapid for you to make out much or to detect anything clearly. A word will be projected between the penciled guide lines you see on the screen. Please report whatever you can make out between the guide lines after each exposure. Do you have any questions?

RESULTS

A logarithmic transformation of the data was performed to reduce heterogeneity. The transformed data were then subjected to an analysis of variance the results of which are presented in Table 2.

It is evident that the main effect for competition is significant while that for frequency is not. This means that words designated as high and low competitors created a discrepancy in mean log recognition time sufficiently large to warrant acceptance at the 5% level of significance. On the other hand, discrepancy between high and low frequency words was not sufficiently large to warrant the assumption of a differential effect. Figure 1 shows the difference between levels of competition and levels of frequency with the *S*s arranged in ascending rank order along the abscissa. The differences are distinct with respect to competition but not so with respect to frequency.

---

[2] We are grateful to Donald Shurtleff for his aid with the design.

TABLE 2

ANALYSIS OF VARIANCE OF MEAN LOG RECOGNITION TIME DATA

Source	*df*	*MS*	*F*
Frequency (F)	1	2,072.07	1.51
Competition (C)	1	7,703.51	5.60*
F × C	1	56.44	<1
Error (words within blocks)	12	1,375.65	
Days (D)	1	876.75	3.07
D × C	1	31.01	<1
D × F	1	765.39	2.68
*S*s	7	721.94	2.52*
S × F	7	554.11	1.94
S × C	7	442.85	1.55
S × D	7	1,484.01	5.19***
S × F × C	7	866.07	3.03*
Error	74	286.03	

* $p = .05$.
*** $p = .001$.

It is also evident that no significant interaction occurred between the two independent variables. The effect of competition proved independent of the frequency of the stimuli.

The remaining sources of variance in the table are concerned with day and *S* effects. The variance for day effect is associated with the two separate experimental sessions, and does not reach significance. However, it is not wholly negative, for while the data indicated that the effect of session is negligible for high frequency stimuli, it is less so for low frequency stimuli. This could be interpreted as an indication that low frequency words showed a practice effect not shared by the high frequency words, a probable conclusion.

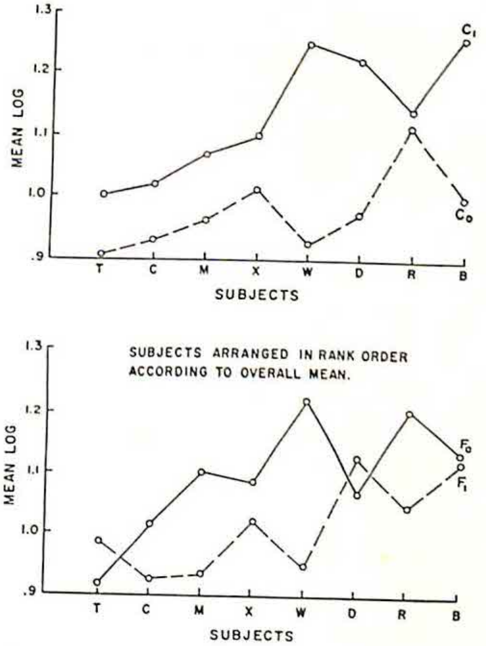


FIG. 1. Mean log response time for 8 *S*s.

Figure 2 depicts the interaction of *S*s and word blocks. The range of individual differences is marked and is responsible for the significant *MS* for *S*s. Figure 2 also shows the mean

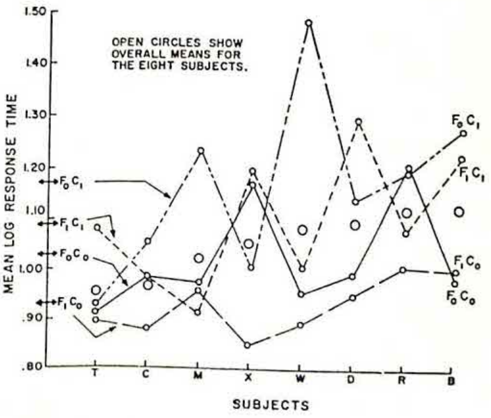


FIG. 2. Mean log X plotted for each *S* for each block of four words.

TABLE 3

Number of High Frequency Responses per Stimulus and Number of High Frequency Responses That Are Competitive Response Matches

Stimuli	Main Variables	
	High Freq. Response	Compet. Response Match
C_1		
lint	128	121
tusk	74	69
snow	38	35
mare	15	4
desk	10	1
beat	85	42
some	38	9
case	26	2
C_0		
wren	17	0
nave	29	0
fief	18	0
drab	18	0
over	18	0
wear	22	0
kind	4	0
book	32	9

recognition time for each word block. The four means progress in the hypothesized order. That is, it was expected that the low frequency high competitor words would be the most difficult to recognize.

Up to this point, comparisons have been between the recognition thresholds of the stimuli predicted to encounter strong or weak competition on the basis of there being in English high frequency response matches for each—a match being defined as a response which has all but a middle letter in common with the stimulus and that letter the same shape as its stimulus counterpart. It remains to be shown that these words did indeed engender strong or weak competition. Table 3 indicates the total number of high frequency matches *S*s reported for each stimulus, and also the total number of high frequency responses. The total number of matches for all eight high competitor words was 283 and that for the low competitor words 9. The separation of stimuli into high and low competitor groups is supported by these results.

Discussion

The present experiment has been an attempt to demonstrate the existence of a competitive or interference effect on recognition thresholds. Such an effect had been previously hypothesized in the theory advanced by Solomon and Postman (1952) to explain the repeatedly confirmed correlation between frequency of word usage and threshold. The present data demonstrate the existence of the competitive effect. We conclude that the availability or unavailability of strong competitive responses affects rapidity of recognition.

Furthermore, under conditions of either high or low competition, there are observed no marked differences of threshold as a result of varying the frequency of prior usage of the word stimuli. The data suggest only a trend toward a frequency effect that does not reach statistical significance. With the reservation that the number of *S*s is small, we can state that in this experiment the frequency of prior word usage of the stimuli was not of itself a significant variable.

This conclusion defies the extensive experimental evidence advanced to support the claim of a stimulus frequency effect on threshold. The evidence is varied, well-controlled, and convincingly presented, and it is in fact not our intention to dispute its accuracy. Instead, we intend to explain the frequency finding in a new way, by reference to what we believe is the more fundamental principle behind it.

What our results suggest is that the frequency finding is an outcome of failure to control competition among prerecognition responses and that when the effect of this competition is controlled the frequency effect markedly diminishes.

It is important to have clearly in mind what is meant here by strong and weak competitors. Strong competitors are prerecognition responses with two characteristics: they are response words with a certain formal or structural similarity to the stimulus and a high frequency of usage. The importance of frequency is, therefore, implicit in the concept of competition, but it is not the frequency of the stimulus that is important but the frequency of the prerecognition responses. Competition acts independently of the frequency of stimuli, while the strength of the competitive response is partly a function of general word frequency. We, therefore, suggest the following revision of the frequency rule: threshold is not a function of the frequency of prior usage of stimuli but of the ability or inability of the stimuli to evoke high frequency competitive responses. These statements are in contrast with those advanced by Solomon and Postman in which both threshold and competition are functions of stimulus frequency.

Why then has a frequency effect been so consistently found? We suggest that the answer lies in a characteristic of any random population of high and low frequency stimuli: the likelihood of there being high frequency matches (strong competitors) for low frequency stimuli is greater than the likelihood of strong competitors for high frequency words. The reason is that the number of suitable form fitting responses is limited for any stimulus and the odds are greater of there being a high frequency match for a low frequency word which already occupies, so to speak, the high frequency slot. The operation of the more fundamental competitive effect has, therefore, lain concealed in the previous experiments on word frequency.

The present experiments do not, however, allow us to conclude that everything noted of the effect of prior usage on threshold is due to competition. The reported results only suggest that much, and perhaps most, of the frequency effect is due to competition. It remains to be seen what other mechanisms may mediate the previously observed frequency effect.

A final word: the present experiments serve to give much needed attention to the formal or structural features of the stimulus, and the effective stimulus fragments, in the determination of threshold. In many experiments an attempt has been made to avoid the structural problem by the use of capitalized stimuli. We doubt that this is ever altogether successful and may only serve to render the exact delineation of the role of structural features more difficult. It has long been known that structural features play an important part in threshold levels, but much remains to be discovered as to the mechanisms behind these threshold differences. Frequency itself may play a role and, as has been here suggested of those complex forms which are words, the effect of frequency may be mediated by competitive processes.

REFERENCES

BRUNER, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

McGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.

McGINNIES, E., COMER, P. B., & LACEY, O. L. Visual recognition threshold as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44**, 65–69.

NEWBIGGING, P. L. The perceptual redintegration of frequent and infrequent words. *Canad. J. Psychol.*, 1961, **15**, 123–131.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition threshold for words. *J. exp. Psychol.*, 1952, **43**, 195–210.

(Early publication received July 6, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 12–21

JUDGMENTS OF WEIGHT AS AFFECTED BY ADAPTATION RANGE, ADAPTATION DURATION, MAGNITUDE OF UNLABELED ANCHOR, AND JUDGMENTAL LANGUAGE[1]

O. J. HARVEY AND DONALD T. CAMPBELL

University of Colorado *Northwestern University*

Ss were first presented a series of 5 weights under varied conditions and then exposed additionally to extreme anchoring stimuli. Half of the 160 Ss made their judgments in ounces (Physical language) and the other half in 1–5 categories (Comparative language). The design under each language was a 2 × 2 × 2 factorial: Narrow and Wide ranges of the standard weights designed to produce common geometric means or adaptations levels; Weaker and Stronger scales, represented by variation in number of preanchor trials; and Lighter and Heavier anchors. All 4 variables significantly affected judgments, indicating the importance of parameters generally not treated in adaptation level theory and suggesting that some of the anchoring effects typically obtained are affected by the judgmental language.

If it is true, as Helson (1947) has implied, that the effects of all previous stimulation upon judgments of a novel stimulus can be summarized as a single value, a weighted mean of previous stimuli, then adaptation levels (ALs) produced by different stimulation histories would have equivalent effects upon the judgment of any novel stimulus.

It is with this question that this study is mainly concerned. The general approach consisted of presenting systematically varied sets of stimulation aimed at producing equivalent ALs and determining their effects upon the judgment of common stimuli. Stimulus variations included adaptation range, anchor magnitude, adaptation duration, and judgmental language.

Adaptation range.—The differences in the ranges of the two sets of weights are depicted in Table 1. If judgments of a stimulus are determined by its relationship to an AL, or geometric mean, then judgments of Weights 3 and 7, as well as of the two anchors, should not be affected significantly by width of stimulus range. It was our assumption, however, that range variation would result in different judgments of the common stimuli and in different reactions to extreme anchors.

Magnitude of anchor.—Will preliminary stimulus series of common ALs but varied ranges be affected similarly by the introduction of more extreme stimuli (anchors)? If range is an important variable, then the effects of an anchor should be greater for the narrow than for the wide scale. To test this assumption, two magnitudes of anchors were used.

Adaptation duration.—Both common sense and results from other studies (e.g., Harvey & Caldwell, 1959; Johnson, 1949; Tresselt, 1947) strongly suggest that an AL based on a large number of experiences with a set of stimuli should be

---

[1] Financial aid for this study, initiated in 1955 when both authors were visiting at Yale, was provided by the Yale University Communications Research Program under the direction of Carl Hovland and supported by a grant from the Rockefeller Foundation, to both of whom we express our thanks. We also wish to thank Jack Donovan for his help in collecting the data and Don Van Ostrand for his help in analyzing them.

TABLE 1

STANDARD AND ANCHORING STIMULI

Wide Range		Narrow Range		Anchors
Stimulus Numbers	Weight (Gm.)	Stimulus Numbers	Weight (Gm.)	Weight (Gm.)
1	305.5			
3	361.9	3	361.9	
		4	393.9	
5	428.7	5	428.7	
		6	466.6	
7	507.8	7	507.8	
9	601.5			
				654.7
				1097.1

less affected by the introduction of extreme anchors than one derived from fewer exposures. This question has not been answered, however, within the framework of AL theory. To answer it, variation was made in the number of times stimuli were presented in the pre-anchor session.

Physical vs. comparative judgment language.—By a physical language we mean to refer to a relatively context-free language, one which is based upon more absolute criteria, such as those used in physics and implied in everyday judgments of such attributes as heaviness in pounds, distances in feet, etc. In this study the physical language is the scale of ounces applied to judgments of lifted weights. This is in contrast to the comparative language generally employed in the method of single stimuli in which such standard terms as "heavy" and "very heavy," or the numerical evaluations of 1–5, for example, imply comparisons which are uninterpretable without knowledge of the context of the comparison.

The physical language, moreover, should be possessed of a greater divisibility and extendability than is the traditional comparative language. Hence the readjustments generally resulting from extreme anchors may be as much a reflection of "vocabulary exhaustion," restriction of the number of usable categories, as of an inability to discriminate among the stimuli. If so, the anchoring effects should be less for an open-ended physical language than for a restricted comparative language. This possibility was tested in the present study by duplication of the experimental conditions with physical and comparative languages, the first time, we believe, that a direct comparison will have been made between anchoring effects on judgments rendered in the traditional categories and judgments made in ounces.

METHOD

Apparatus.—The experimental stimuli were small cylindrical glass jars, of the same size and filled with varying amounts of lead shot, sand, and cotton in such a way that the balance of the weights was sufficiently similar that minimal gravitational cues were provided for identification of particular weights. Judgments of the weights, instead of being announced orally, were recorded by *S*s on a 10-key electric adding machine (placed on a table that also supported the arm of *S* not involved in lifting weights). This way of communicating judgments resulted from the uncommon method of collecting data from more than 1 *S* at a time, which, in order to prevent social influence, necessitated that *S*s not make their judgments aloud. By means of four 5 × 5 ft. plywood partitions and automatic timing devices 4 *S*s served simultaneously. The partitions extended diagonally from the corners of a 5 × 5 ft. square in which *E* sat and controlled the presentation of the weights. The *S*s sat outside the square, 1 against each side, with their right arms extending into the inside through a cloth covered opening in each side of the square. Attached to inside walls of the square at appropriate places were arm rests cushioned with foam rubber on which *S*s reclined their arms, palms upward, between trials. When lifts were being made, the arm was free of the rest, the lift being made from the shoulder for greater accuracy of judgments (Oberlin, 1936). Each *S* was presented a stimulus every 30 sec., following a ready signal represented by a brief flash of a small red bulb on the partition in front of him and activated by electrically sensitized contact points on a rotating timer-wheel that was driven by a 1-rpm synchronous motor. Eight contact points were attached to the timer at equal distances, so that the time interval between the presentation of a stimulus to each succeeding *S* was 7½ sec. and the time between the trials of a given *S* was 30 sec. Immediately following the ready signal, E, seated on a swivel chair inside the enclosed square placed the weight flatly on the palm of the appropriate S who then had 7½ sec. in which to lift the weight and form a judgment of it.

Procedure.—The study was carried out as two experiments, one involving use of the *Physical* judgmental language and the other, a replication, using the *Comparative* judgmental categories. The design in each of the language variations was a 2 × 2 × 2 factorial: *Narrow* and *Wide* range of the sets of stimuli entering into the formation of presumably equivalent ALs; *Weaker* and *Stronger* scales or varied adaptation duration; and *Lighter* and *Heavier* anchoring stimuli, with blocks of trials cutting across each condition.

In the physical language condition judgments of heaviness of the weights were recorded in ounces, while in the comparative language variation judgments were made in the numerical categories of 1, 2, 3, 4, and 5. Stability duration of the AL was varied in terms of the number of preanchor presentations of the standard stimuli: 10 presentations for each of the five weights in a series for the Weaker scale and 25 presentations of each of the five weights in the Stronger scale. The differences between the Narrow and Wide sets of stimuli employed to produce what was presumed to be equivalent ALs can be seen from Table 1. For anchors, the Light anchor of 654.7 gm. was only one log-unit step above the upper limit of the wide series, while the Heavy anchor of 1097.1 gm. was seven units above this value. In devising the two sets of stimuli calculated to produce the same AL, we deliberately ignored the "d" factor (Helson, 1947, p. 6) because to have applied it here would have assumed a differential effect of the narrow and wide range conditions which instead was one of the major effects we were interested in demonstrating.

The experiment was divided into two sessions, separated by a rest interval of 5 min. during which conversation between Ss was not permitted. In the first session, aimed at the formation of scales and ALs of differing stability, Ss were exposed to five standard weights of either the wide or narrow range for either 10 or 25 presentations per weight which were judged in either ounces or categories 1–5. In the second session Ss were exposed to either the lighter or heavier anchor for 50 trials along with 10 more presentations of each of the standard weights they had judged in the preanchor session. The anchor weight was always preceded by one of the standard stimuli. In both the scale formation and anchor sessions, the presentation of the standard stimuli was randomized within each round of five weights for each group of 4 Ss.

To avoid confounding effects of labeling with the physical extremity of the anchor stimulus, the anchors were not labeled in this study, unlike most studies involving anchoring effects, nor were any cues given to Ss by E to indicate that a heavier weight was being introduced. Here Ss were free to apply their judgmental categories to the total range of stimuli, including both the standard and anchor weights in whatever way they deemed appropriate without the usual instructional restriction that the extreme categories should be applied to the extreme, anchoring stimulus.

The actual procedure was as follows: After Ss had been properly seated and shown how to use the adding machines to record their judgments, they were advised that their task was to judge the heaviness of a number of weights from lifting them. In sequence, they were instructed that a weight would be placed in the palm of their hand each time the bulb flashed in front of them. The weight was to be lifted, from the shoulder, for 7½ sec. and then the judgment of its heaviness recorded on the machine. Depending on whether the judgmental language condition was physical or comparative, judgments were recorded in either ounces or categories, 1 for the lightest and 5 for the heaviest weight in the latter case.

To prevent Ss from using the noises from the other adding machines as cues for their own judgments, they were erroneously told that they would receive different weights. Social influences were further minimized by not allowing conversation during the experimental sessions.

Subjects.—The Ss were 160 undergraduate students enrolled at Vanderbilt University during the fall of 1957, 10 in each of the 16 conditions of this study. The data from 20 Ss who participated in a pilot study at Yale University are not included.

RESULTS AND DISCUSSION

Effect of adaptation range on preanchor judgments.—Table 1 shows that Weights 3, 5, and 7 were common to both the narrow and wide range of stimulus weights. If the AL in this case may be represented by a geometric mean, then both ranges of weights should have produced an equivalent AL and Weight 5 should not have been judged significantly differently under the two scale widths. If, on the other hand, the range within which the common weights are embedded is important, then Weights 3 and 7 should be judged as lighter and heavier, respectively, under the narrow than under the wide ranges, appearing more extreme when representing the end points of the scale than when being surpassed at both the upper and lower limits by more extreme stimuli (Weights 1 and 9). In Table 2 are presented the mean judgments of Weights 3, 5, and 7 in the preanchor session for both categories and ounces languages.

The data from which the results in Table 2 were summarized were treated by analyses of variance (Lindquist, 1956), a separate analysis for each of Weights 3, 5, and 7 under the narrow and wide ranges, and weaker and stronger scales across blocks of five judgments in the preanchor phase, for a total of 12 analyses for the three common weights. Since these analyses were based on preanchor judgments, lighter and heavier anchor groups were combined into a single group. But because the weaker and stronger scales were based on different frequencies of preanchor trials (10 vs. 25 trials per weight) separate analyses were necessary for the variations in scale strength.

In confirmation of the intent of the design, judgments of Weight 5, the numerical value of which was presumed to represent the AL for both the narrow and wide ranges, were affected almost not at all by variation in adaptation range for either language in either the stronger or weaker scale (average $F = .68$). Judgments of Weights 3 and 7, on the other hand, were significantly influenced ($p < .01$) by the difference in range when rendered in the 1–5 categories (F for strong scale $= 9.00$, and for weak scale $= 11.33$), with Weight 3 being judged as lighter and Weight 7 as heavier in the narrow presentation, as was predicted. When made in ounces, however, judgments of Weights 3 and 7 in both the strong and weak conditions were not significantly affected by range differences (average $F = 1.36$).

TABLE 2

Mean Judgments of Common Weights 3, 5, and 7 in the Preanchor Session

Weights	Strong Scale[a]		Weak Scale[b]	
	Narrow Range	Wide Range	Narrow Range	Wide Range
	Categories Language			
3	2.05	2.50	2.00	2.40
5	3.30	3.35	3.35	3.45
7	4.45	4.10	4.55	4.15
	Ounces Language			
3	12.70	9.20	10.05	11.05
5	17.40	13.35	15.50	16.10
7	23.80	18.90	21.80	24.15

[a] 25 Preanchor judgments per weight.
[b] 10 Preanchor judgments per weight.

The differential effect of range for the two judgmental languages may be highlighted by a more direct comparison, the difference between differences of judgments of Weights 3 and 7 under the narrow and wide ranges in the preanchor series. For each *S* a single measure of spread was computed by subtracting the mean judgment of Weight 3 from the mean judgment of Weight 7. Since there were no significant differences in these spread measures between the weak and strong conditions, these two conditions were collapsed, providing 40 *S*s under each of the narrow and wide ranges. The difference between these (7–3) span measures provided a t of 7.62 ($p < .01$) for the category judgments and a t of only .10 for the ounces judgments (with the tendency for a wider span under the narrow range for both judgmental languages). The difference between these differences in range effects for the two languages yielded a t of 5.32 (using 1.414 as the standard error of the difference between two t's under large sample conditions), significant at better than the .001 level. Thus the effect of adaptation range on judgments of Weights 3 and 7 appears nil when judgments were made in ounces, but marked when the judgments were given in the discontinuous categories.

The fact that differences in stimulus ranges significantly affected categories judgments of Weights 3 and 7 cannot be explained by our omission of the "d" factor in determining the AL. Had the "d" factor (Helson, 1947, p. 6) been included, the AL for the narrow range would be 401 gm. and the AL for the wide range would be 373 gm. If AL were the major determinant of judgments, then all three weights, 3, 5, and 7, should have been judged heavier in the wide range condition because of the lower AL value. This is not what obtains. Instead, in the wide range Weight 3 was judged significantly heavier and Weight 7 significantly lighter, with no range differences for Weight 5, all of which is more compatible with range determination than AL influence.

Effects of anchors on judgments of Weights 3, 5, and 7.—In Table 3 are presented the means and t ratios based on the differences between judgments of Weights 3, 5, and 7 in the preanchor and anchor sessions

TABLE 3

MEANS AND t RATIOS BASED ON CHANGES IN JUDGMENTS OF WEIGHTS 3, 5, AND 7 OCCURRING IN THE ANCHOR SESSION

Cond.	Categories					
	Weight 3		Weight 5		Weight 7	
	$\bar{x}$ Diff.	t[a]	$\bar{x}$ Diff.	t	$\bar{x}$ Diff.	t
NSL	.84	6.70	1.05	9.39	1.40	10.32
NWeL	.71	7.14	1.22	9.68	1.51	8.16
NSH	.66	3.84	1.41	4.77	1.86	6.73
NWeH	.76	4.27	1.58	10.46	2.14	12.09
WiSL	.62	7.60	.55	6.62	.71	8.31
WiWL	.51	2.62	.80	4.44	.98	4.32
WiSH	1.12	7.71	1.55	11.65	1.67	12.36
WiWeH	.91	5.52	1.33	19.97	1.72	20.07
	Ounces					
NSL	.15	.15	.74	.68	.44	.33
NWeL	1.00	1.41	2.59	2.93	2.95	2.44
NSH	6.03	3.34	8.86	3.60	11.38	3.96
NWeH	2.15	2.12	6.82	4.35	7.63	4.78
WiSL	1.75	4.18	2.76	5.01	3.65	5.54
WiWL	1.53	1.47	−.20	.18	.21	.73
WiSH	2.01	3.15	3.43	3.90	5.55	3.37
WiWeH	3.13	5.94	4.14	4.80	5.84	4.40

Note.—N for each condition = 10.

Abbreviations for experimental conditions: N = narrow adaptation range; Wi = wide adaptation range; We = weak preanchor scale (10 trials per weight); S = strong preanchor scale (25 trials per weight); L = lighter anchor; and H = heavier anchor.

[a] For $df = 9$, t for $p_{.05} = 2.26$, for $p_{.01} = 3.25$.

for each of the 16 experimental conditions. In 47 of the 48 cases introduction of the anchors (heavier weights) resulted in the standard weights being judged as lighter than the preanchor evaluation of them, significantly so in all conditions under the categories treatment and in most of the conditions under the ounces language. The heavier anchor produced significantly greater judgmental change on each of the three weights under both languages than did the lighter anchor ($p < .001$).

In anticipation of a more statistically precise and appropriate comparison which will be presented later, it should be noted that the anchoring effects were consistently more pronounced when the categories language was used than when the ounces distinctions were made. Of the 24 differences compared under the categories language, 23 were significant at less than the .01 and the other one at less than the .05 level. On the other hand, only 13 of the compared differences in the ounces language were significant at less than the .01 level while 8 failed to attain the .05 level of significance. The results in Table 3, however, demonstrate clearly for both the ounces and categories judgments one of the most general and basic effects of stimulus context, viz., the relativity of experienced stimulus intensity to the background or membership character of the stimulation.

Adaptation range and effects of anchors on scale shift.—The shift toward judging Weights 3, 5, and 7 as lighter following introduction of the anchors was expected to be greater for the narrower than for the wider adaptation range. F tests based upon the data summarized in Table 3 showed this to be only equivocally the case. On the total shift in judgments of the three weights combined, variation in scale width exercised a significant effect, with judgments in the narrower scale changing more than those in the wider scale under both the ounces ($F = 4.00$, $p < .05$) and categories ($F = 11.60$, $p < .01$) language. With change treated separately for each of the three weights, however, scale width differences failed to influence significantly judgments of Weight 3 under both languages ($p < .10$) as well as of Weight 7 under the ounces treatment ($p < .10$). Hence adaptation range produced anticipated differences only for judgments of Weights 5 and 7 in the categories language ($p < .05$ and $< .01$, respectively) and of Weight 5 in the ounces condition ($p < .05$). Further, the magnitude of anchor failed to interact significantly with scale width in effecting judgmental change.

Absence of effects of scale breadth on shifts in judgments of Weight 3 under the categories language may be the result of the remoteness of this weight from the anchors, a possibility implied by the fact that the anchoring effects (toward lighter judgments) on the standard weights progressively decreased with increased remoteness of the weights from the anchor. This possibility does not apply, however, for the irregular effects of scale width under the ounces treatment for not only were judgments of Weight 3 unaffected when given in ounces but judgments of Weight 7, close to the anchor, also were unaffected. It is probable, therefore, that the lack of scale width effects under the ounces treatment stems from the greater open-endedness and elasticity of this judgmental framework.

Effect of adaptation duration or strength of preanchor scale on judgment shift.—From Table 2 it may be noted

TABLE 4

MEANS AND t RATIOS BASED ON THE DIFFERENCES BETWEEN THE FIRST FIVE AND LAST FIVE JUDGMENTS OF WEIGHTS 3, 5, AND 7, IN THE ANCHOR SESSION

Language	Weight 3		Weight 5		Weight 7	
	$\bar{x}$ Diff.	t[a]	x Diff.	t	$\bar{x}$ Diff.	t
Categories	.12	2.46	.13	2.23	−.10	−1.28
Ounces	.59	2.00	.17	2.59	−.47	−1.46

[a] t for $p_{.05}$ at $df = 78 = 2.00$.

that 25 preanchor trials per standard weight rendered the concept of the weight no more stable or resistant to influence by the anchor than did only 10 exposures to each weight. Analyses of variance based on the scale changes showed no significant difference between the two practice frequencies for any of the three weights in either language treatment (largest $F = .68$). Nor did adaptation duration interact significantly with either width of scale or magnitude of anchor.

Effect of repetition of anchor on shift of scale.—This effect was tested by t ratios based on the mean difference in amount of shift between the first five and last five judgments of the common weights 3, 5, and 7 in the anchor session, collapsed across scale width, strength, and anchor. From the results in Table 4, it can be seen that while judgments of Weight 7 under neither language was differentially affected by the temporal variation of the anchors, judgments of Weights 3 and 5 were significantly more depressed by the first five than by the last five exposures to the anchor weights. Anchor trial blocks, however, did not interact significantly with scale strength, width, or anchor, as determined by separate t tests under each of these conditions.

The greater effect of the earlier experiences of the anchor is contrary to Helson's theory when the trials in both the preanchor and anchor sessions are all regarded as part of the "specious present." The suggested modification of Helson to weight recent judgments more heavily than remote ones would predict a still sharper trend in the direction opposite to that obtained here. The simplest, and perhaps most plausible, explanation of the greater effects of the earlier anchor experiences is that early in the anchor phase S tended more to confuse the anchor with the scale weights, a tendency which dissipated as further experience allowed him to discriminate more accurately between the anchor and the other weights.

Judgmental language and effects of anchor on shift of scale.—In order to make a more direct comparison of the anchoring effects on the judgments made in ounces and those rendered in categories it was first necessary to convert the shifts to some sort of common denominator. This was done by transforming the judgmental changes into z's, which in this case were represented by the difference between the means of an S's pre- and postanchor judgments of each of the three common weights, 3, 5, and 7, divided by the SD for the preanchor judgments in eight conditions combined under each of the language treatments. The SDs were thus derived separately for judgments in the two languages, and the mean change in an S's judgment subsequently was expressed in units of the appropriate SD represented by his change.

In Table 5 is presented a comparison of the shifts in scales produced by the anchors under the two judgmental language treatments. For each of the three common weights, 3, 5, and 7, anchoring effects were significantly greater for the judgments made in categories than those given in ounces, something that has been implied several times in our results reported so far. Thus, while demonstrating that the anchoring effects long observed in the method of single stimuli obtain both for more restricted and less restricted judgmental language, this study has also shown that the anchoring effects are accentuated when *S* is faced with the necessity of fitting extreme (anchoring) stimuli into categories numerically sufficient for only less extreme (standard) stimuli. Since the extreme stimulus must be included in the more closed evaluative system, it is not surprising that an alteration in the application of the same categories to preanchor stimuli is the outcome. Given a more unbounded scale, as in the ounces judgments, *S* is not forced to change his preanchor evaluations because of his freedom to employ new, more extreme, evaluations to encompass the more extreme stimuli. The changes due to the introduction of anchors under the usual method of single stimuli are shown to be due to at least two factors, one presumably due to the restricted judgment language, one to a shift in adaptation level or judgment base. Presumably the latter obtains more heavily in the cases of discriminations made in ounces.

TABLE 5

JUDGMENTAL LANGUAGE AND MEAN SHIFT IN JUDGMENTS OF COMMON WEIGHTS 3, 5, AND 7

Weight	Categories	Ounces	Diff.	t^a
3	1.688	.428	1.260	2.703
5	2.618	.426	2.192	4.810
7	2.931	.369	2.562	5.492

Note.—Based on *z* scores.

[a] t for $p_{.05}$ at $df = 78 = 2.00$.

The results presented thus far pertain to effects of the anchor upon scale shift as a function of the several independent variables. Also of theoretical significance, however, is the effect of the treatments upon judgments of the anchors themselves, extractable in this study because of the unlabeled nature of the anchors. The mean judgments of the anchors under each of the main conditions are presented in Table 6. From separate analyses of variance for the two languages, strength of scale and magnitude of anchors yielded a significant effect on judgments of the anchors under both languages while width of scale showed no significant effect on anchor judgments in either. In the ounces language the *F* for scale strength was 12.71 ($p < .01$) and for anchor magnitude was 108.16 ($p < .001$). The respective values in the categories treatment were 9.09 ($p < .01$) and 37.33 ($p < .001$).

The finding that strength of scale did not affect the influence of the anchors on the scale but significantly affected judgment of the anchors themselves parallels the results obtained by Harvey and Caldwell (1959) in a related study. This outcome is doubtlessly consistent with various theoretical interpretations, rela-

TABLE 6

MEAN JUDGMENTS OF ANCHORS, COLLAPSED ACROSS TRIAL BLOCKS

Judgmental Language	Strong				Weak			
	Light Anchor		Heavy Anchor		Light Anchor		Heavy Anchor	
	Narrow	Wide	Narrow	Wide	Narrow	Wide	Narrow	Wide
Ounces	28.19	18.08	58.16	46.44	25.14	47.78	66.58	47.55
Categories	4.82	4.60	4.85	4.95	4.52	4.56	4.78	4.40

tive effort and fatigue being among the more immediately suggested possibilities. If this were the explanatory variable, however, the heavier anchor should have been experienced progressively heavier with repetition of the anchor. Yet notably absent for both ounces and categories judgments was any effect approaching significance of differences in trial blocks of the anchors, either as a main effect or in interaction. Hence other explanations have to be invoked. We prefer the possibility that stronger and better defined concepts (of heaviness, of social objects, or of whatever) result in a narrower latitude or defining limits of the concept with the result that the same stimulus phenomenologically would be more distant from it than from a weaker concept of wider and less definitive limits. In support of this possibility is the common observation that it is the person with the strong and narrow concepts who is more sensitive to events that deviate from his concepts and who is more likely to exaggerate the deviation. Thus the more opposed individuals were to the use of alcohol, the greater was their tendency to exaggerate the "wetness" of a communication which in reality was only slightly favorable to alcohol (Hovland, Harvey, & Sherif, 1957).

Width of scale and anchor magnitude interacted significantly in judgments of the anchors for both judgmental languages (F for ounces $= 19.33$, $p < .001$; F for categories $= 6.373$, $p < .02$). But the direction of the interaction in the two cases was opposite. In the ounces language the lighter anchor was judged significantly heavier in relation to the wider scale while the heavier anchor was evaluated as being significantly heavier in relation to the narrower scale. Just the reverse was the case for the categories judgments. Strength (based on variations in number of preanchor trials) interacted significantly with anchor for ounces judgments only ($F = 4.07$, $p < .05$), with the lighter anchor being judged significantly heavier in relation to the weaker scale, but with no significant difference between scale strength in judgment of the heavier anchor. And in both languages a significant interaction occurred among strength, width, and anchor (F for ounces $= 17.41$, $p < .01$; F for categories $= 5.80$, $p < .05$).

Effects of anchors on fineness of discrimination.—One of the traditional interests in the method of single stimuli, particularly as summarized by Volkmann (1951), is in the effects of the anchor on the fineness of discrimination resulting from the experimental treatments. Two measures of such effects were employed in this study: changes in *SD*s of the judgments following introduction of the anchor (closer to the more traditional measure), and changes in coefficients of correlation between the actual weight and judgments of that weight (see Campbell, Lewis, & Hunt, 1958, for a similar analysis). Thus entering into the latter measure of change in scale discrimination were two correlation coefficients (corrected for small number of intervals in the case of the category judgments): one based on preanchor judgments of the standard stimuli and the other based on judgments of the same stimuli after introduction of the anchor.

The effect of judgmental language on changes in discrimination was measured by a t test based on a conversion of the differences between r's under the two conditions to normal deviates. The normal deviates were determined by dividing the difference between the correlation of each S's preanchor and anchor judgments by the SD of all preanchor correlations under both categories and ounces conditions. (In the stronger scale treatment the correlations entering into the overall SD were from the last 50 of the total 125 preanchor judgments.) The difference in change in scale discrimination between the two judgmental languages yielded by this analysis did not approach significance (mean difference = .391; t = .744).

With the establishment of no difference between judgmental languages, the r's for both sessions were transformed to Fisher z's and treated by an analysis of variance for the effects of strength and width of scale and magnitude of anchor on alterations in scale discrimination. Significantly greater loss in discrimination occurred in the narrow than the wide scale ($F = 92.27$; $p < .001$) and for the heavier than for the lighter anchor ($F = 5.28$; $p < .05$). Strength of scale had only a very negligible effect on changes in discrimination as did the interaction of the variables (F for total interaction = .328).

In each of the 16 conditions the introduction of the anchor reduced judgmental variability for a large majority of the *S*s, with 96% of the total sample being so affected. The frequency of this reduction, according to the sign test (Dixon & Massey, 1951), was statistically significant for all eight conditions under the categories language and for six of the conditions in the ounces treatment. In Table 7 is presented the mean difference in magnitude of this variability change between scale widths, strengths, and anchors. In the categories language a trend towards greater reduction in variability occurred (only at the .10 level) for the narrow than the wide scale; significantly greater reduction in variability occurred for the weak than the strong scale and for the heavy than the light anchor ($p < .05$). In the ounces treatment, however, only anchor magnitude had an appreciable effect on reduction of variability, with the heavy anchor exercising a more stabilizing influence on judgments than the light anchor (only at the .10 level).

TABLE 7

CHANGES IN JUDGMENTAL VARIABILITY (*SD*) FROM INTRODUCTION OF THE ANCHOR

Comparisons	Ounces		Categories	
	$\bar{x}$ Diff.	t	$\bar{x}$ Diff.	t
Narrow-Wide scale	.64	.947	.18	1.784
Weak-Strong scale	.62	1.092	.17	2.240*
Heavy-Light anchor	1.14	1.823	.14	2.075*

* $p < .05$.

REFERENCES

CAMPBELL, D. T., LEWIS, N. A., & HUNT, W. A. Context effects with judgmental language that is absolute, extensive and extra-experimentally anchored. *J. exp. Psychol.*, 1958, **55**, 220–228.

DIXON, W. J., & MASSEY, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.

HARVEY, O. J., & CALDWELL, D. F. Assimilation and contrast phenomena in response to environmental variation. *J. Pers.*, 1959, **27**, 125–135.

HELSON, H. Adaptation-level as frame of reference for prediction of Psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

JOHNSON, D. M. Learning function for a change in scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–860.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

OBERLIN, K. W. Variation in intensive sensitivity to lifted weights. *J. exp. Psychol.*, 1936, **19**, 438–455.

TRESSELT, M. E. The influence of the amount of practice upon the formation of scale of judgment. *J. exp. Psychol.*, 1947, **37**, 251–260.

VOLKMANN, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 273–294.

(Received January 8, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 22–29

CONCEPT SHIFTS AND VERBAL BEHAVIOR [1]

ROY LACHMAN AND JOYCE A. SANDERS

Hollins College

2 experiments on concept shifts were conducted to test a theoretical model of thinking. The model assumed that mediating responses follow the laws of overt verbal responses. Ratings of verbal labels for pairs of stimulus values confirmed an a priori shift classification of similar, neutral, and opposite. Transfer effects for the 3 types of shifts were compared on successive concept shifts. A quadratic transfer function was obtained for the similar, neutral, and opposite shift dimension. Inflection of the curve occurred at the neutral position. A negative displacement of the transfer function occurred after the 1st shift. The results were judged as more consistent with Bugelski and Cadwallader's reformulation of Osgood's transfer laws than with Osgood's original formulations.

The present paper is concerned with discovering the effects of adding another kind of shift in a reversal-nonreversal learning situation (Kendler & Kendler, 1962). A reversal shift involves reversing the reinforcement contingencies for two values of a relevant stimulus dimension; for example, reinforcement is changed from black (+) and white (−) to black (−) and white (+). A nonreversal shift changes the relevant dimensions, for example, black (+) and size irrelevant is shifted to large (+) and brightness irrelevant. By adding a third stimulus value to the dimensions of size and brightness (i.e., medium size and gray), it is possible to institute an intradimensional shift which involves an end and central point of a stimulus dimension, for example, reinforcement is shifted from black (+) to gray (+). A reversal shift, in contrast, involves an intradimensional shift of the two polar points of a stimulus dimension.

In terms of the concepts described: black-gray, black-large, and black-white, the three types of shifts referred to above may be viewed as tests for transfer across semantically *similar* (S), *neutral* (N), and *opposite* (O) concepts. Validation of the classificatory scheme S, N, O was undertaken by requiring *S*s to make judgments of the verbal labels for the successively correct concepts in terms of their similarity, neutrality, or opposition. The ratings were also used to permit comparisons between transfer effects in concept shifts and verbal learning data related to Osgood's transfer surface (Bugelski & Cadwallader, 1956; Osgood, 1946, 1949).

Two studies are presented. The first investigates relative postshift acquisition rates for the three types of shifts and *S*'s S-N-O ratings of the verbal designations for the relevant stimulus values. In both experiments, the effects of overt verbalization were compared to implicit verbal responding, and the consequences of informing *S* that reinforcement is being shifted were also investigated.

EXPERIMENT I

Method

Design.—A 3 × 2 × 2 × 2 factorial design was used involving concept shifts (S, N, and O), instructions (informed and not informed about the shift), articulation (overt and

---

[1] This research was supported by Grant M-5729(A) from the National Institutes of Health, Public Health Service.

covert verbal behavior), and relevant stimulus dimensions (size and brightness).

Subjects.—The *S*s were 96 female students at Hollins College with an age range of 18 to 20 yr. Two additional *S*s were discarded due to reinforcement programing errors.

Apparatus.—An Esterline-Angus chart drive was mounted in a 12.5 in. high × 9 in. long black enamel container with a 10° backward tilt. A transparent aperture, 5 in. long × 2.5 in. high, was located in the center of the container. Two inches below, a response box was mounted. The forward surface of the response box was 7 in. long × 4.25 in. high and was mounted with a 30° tilt. Two pilot lamps, .5 in. diameter, were mounted on the forward surface of the response box, 1 in. from the top. The lamps were set 3 in. apart, one was marked "correct" and the other, "wrong." Three push-button switches were mounted 2.5 in. apart and located 1 in. below the pilot lamps.

One hundred sixty successive sets of three stimuli (i.e., three that were viewed simultaneously) were mounted on a 6-in. strip chart. The strip chart was transported down the stimulus aperture by the chart drive at the rate of 12 in/min. The continuous motion of the chart produced pacing of the choice response. The successive sets of stimuli were mounted 2.25 in. apart, vertically, and remained in full view for 10 sec. The *S* indicated her choice response by pressing one of the three button switches; this illuminated the appropriate pilot light for 1 sec. through a relay holding circuit. The reinforcement lamps were programed manually for each set of stimuli.

Stimuli.—The stimuli consisted of three values for each of five dimensions: geometric form (circle, square, triangle), background color (red, brown, green), number of forms (one, two, three), size of the forms (large, medium, small), and brightness of the forms (black, gray, white). Of the 7776 (6^5) possible arrangements of successive sets of three stimuli, 160 were randomly selected with the following limitations: (*a*) No arrangement of any dimension (e.g., black on the left, white in the middle, and gray on the right) repeated on two successive trials. (*b*) No runs (e.g., black and large appearing together) were repeated for more than four successive trials.

The geometric forms were punched into .75 × 1.5 in. strips of Mystik plastic coated red, brown, or green cloth tape. Slips of Color-Vu black, No. 196 (gray), and white paper were patched behind the cloth tape to produce the brightness levels of the geometric forms. The stimuli strips were then mounted on the chart in sets of three that contained each of the three values of all five dimensions.

Procedure.—Each *S* was individually seated before the apparatus, and told to put on a set of headphones. Nine minutes of taped instructions were transmitted. The nature of the task was narrated in detail. The notion of concept attainment was described with emphasis on the discovery of a common characteristic.

Preliminary training was coordinated with the taped instructions. The pretraining was conducted with stimuli consisting of six successive sets of three letters: A b *c*; c B *a*; etc. The *S* was informed that capital letter was correct; she was instructed to push the button under the capital letter in the first stimulus set and to observe that the lamp marked correct was illuminated. Instructions for the second set of stimuli were to push the button under something other than a capital letter and to observe that the wrong lamp was illuminated. The *S*, then, was instructed to practice on the remaining four sets of sample stimuli but she was to state the characteristic of the stimulus she was selecting before pushing the appropriate button. You are to respond, *S* was informed, by pressing only one of the buttons when the stimuli start to move out of your visual field.

When the pretraining was completed, a set of the experimental stimuli was lowered into the aperture. Each dimension and the values for each dimension were described. The *S* was then directed to remove her headphones.

Half of the *S*s, the *instructed* section, received the following information, "When you are able to light the correct lamp consistently, I will inform you that the problem is being changed and that a different object is correct." The other half of the *S*s, *noninstructed*, was not provided with this information.

All *S*s were also divided as follows. The *covert* section was told, "There is one thing that was not mentioned: you are to say the characteristic of the object to yourself, not out loud." The instructions for the pretraining session to vocalize overtly the characteristic of the stimulus to which *S* was responding, were not countermanded for the *overt* condition.

Each *S*, at the outset, was randomly assigned to one of three shift groups: S, N, or O. Initial training was conducted until *S* achieved a criterion of 10 successive correct responses. At that point, reinforcement was shifted to a different stimulus value and training continued until *S* achieved a second criterion of 10 successive correct responses.

Group S was subdivided and trained with the following pairs of initial and shifted positive stimulus values: gray-white, gray-black, medium-small, medium-large. Group N was subdivided and trained with black-large, black-small, white-large, white-small, large-black, large-white, small-black, and small-white successively positive. Group O, likewise, was subdivided and trained with black-white, white-black, small-large, large-small successively positive. In this fashion counterbalancing was achieved and all *S*s were shifted either to black, white, large, or small.

The *S*s that did not achieve the criterion in 50 trials were told the correct stimulus value, allowed to perform the 10 criterion responses, and then shifted to the second stimulus value. A score of 50 was arbitrarily assigned to these *S*s.

At the termination of training, *S* was handed a rating sheet containing 49 word pairs, including the names of the relevant stimulus values. The *S* was instructed to sort the word pairs into categories of very similar, similar, neutral, and opposite. Before leaving, *S* was directed to sign an honor system pledge committing her to silence concerning the experiment.

Results

The dependent variable employed was the number of trials, excluding criterion trials, required to achieve 10 successive correct responses. Stimulus dimensions produced the only significant effect during the initial training, $F(1,72) = 5.77, p < .05$. Size was the more difficult dimension with a mean of 26.31, the mean for brightness was 17.17. The means for large, medium, and small were not significantly different nor were the means for black, gray, and white (Duncan's range test, $\alpha = .05$). The means for Groups S, N, and O were, respectively, 20.28, 24.71, and 20.22 ($F < 1$ for shifts). Eighteen *S*s (19% of the sample) failed to reach the criterion in 50 trials and were treated in the manner described above before the shift was administered.

Treatment means for the postshift phase of training are presented in Table 1. Five *S*s (5% of the sample) were truncated at 50 trials. Trials to criterion for *S*s in the neutral-noninstructed condition were counted from the first incorrect response following the shift. The analysis of variance for these data is presented in Table 2. Shifts produced the only significant treatment effect. A trend analysis (Grant, 1956) yielded a significant quadratic component for shifts. Although the assumption of equal intervals cannot reasonably be

TABLE 1

MEAN TRIALS TO CRITERION FOR TREATMENTS DURING POSTSHIFT LEARNING

Treatment	Mean Trials to Criterion
Similar	4.97
Neutral	18.47
Opposite	4.75
Covert	7.56
Overt	11.23
Instructed	7.75
Noninstructed	11.04
Brightness	8.83
Size	9.96

TABLE 2

ANALYSIS OF VARIANCE OF TRIALS TO CRITERION FOR POSTSHIFT LEARNING

Source of Variance	*df*	*F*
Shifts (S)	2	16.91***
Linear	1	.01
Quadratic	1	33.81***
Articulation (A)	1	2.76
Instructions (I)	1	2.23
Stimulus dimensions (D)	1	.26
S × A	2	.42
S × I	2	.20
S × D	2	1.43
A × I	1	.28
A × D	1	1.28
D × I	1	.20
S × A × I	2	1.61
S × A × D	2	.18
S × I × D	2	1.56
A × I × D	1	1.37
S × A × I × D	2	.29
Residual (*MS*)	72	(116.86)

*** $p < .001$.

met by the techniques employed, the multipliers for the quadratic component are identical to those used in an orthogonal comparison between the mean for Group N and the means for Groups S and O. The trend analysis implies nothing more than an initial attempt to explore relationships between the present data and the implications of previous research on verbal similarity.

Word-pair rating data were consistent with the experimental classification of S, N, and O. The S concepts (medium-small, medium-large, gray-black, gray-white) were rated similar by 78 to 81% of the *S*s and very similar by 1 to 6% of the *S*s. The N concepts (small-black, small-white, large-black, large-white) were rated neutral by 92 to 97% of the *S*s. The O concepts (black-white, large-small) were rated opposite by 96 and 98% of the *S*s, respectively.

EXPERIMENT II

Positive transfer of training occurred for all three types of shifts in Exp. I. The extensive instructions and pretraining may have failed to produce an asymptotic level of acquisition for task training factors such as learning sets. Negative transfer for one or more types of shift, thus, may have been obscured. Perhaps, the only way to eliminate task learning factors is to have *S* practice on the actual experimental task. This may be accomplished in the present investigation by administering all three types of shifts to each *S* in a counterbalanced order. Considerable methodological difficulty is involved with such a procedure: techniques are not available for a comprehensive statistical evaluation of the data. An independent statistical analysis, however, can be performed on each of the three successive shifts and some understanding of differential transfer effects may be achieved.

Method

Subjects.—Eighty of the *S*s used in Exp. I were subdivided into five orders of shift sequences: S, O, N; S, N, O; N, O, S; O, N, S; and O, S, N. The sixth shift sequence, N, S, O, required 16 additional *S*s who were run concurrently with the *S*s of the first experiment. Thus, 96 *S*s were used and the first of the three shifts in Exp. II employs the data of 80 *S*s of Exp. I. Shifts were administered as soon as previous learning was completed. The *S*s employed solely in Exp. II were run simultaneously with the *S*s serving in both experiments.

Procedure.—Since Exp. II is partially a continuation of the first experiment, the apparatus, stimuli, pretraining, and instructions are identical. The innovation involved administering two additional shifts to the appropriate *S*s.

Type S shifts occurring in the N, O, S and O, S, N sequences and Type N in the N, S, O and O, N, S sequences required counterbalancing of the gray and medium stimulus values. All remaining shifts involved a counterbalanced switch in reinforcement between the black and white, or large and small stimulus values. Complete counterbalancing of the size and brightness dimensions was thus achieved. The two types of shift instruction and the overt and covert conditions of articulation were maintained throughout the four phases of training: initial learning and Shifts 1, 2, and 3.

Results

The response measure used was differences in trials to criterion for all contiguous pairs of training phases. That is, each *S*'s score on Shift 1 was subtracted from her initial learning score, her score on Shift 2 was subtracted from the number of trials to criterion on Shift 1, and trials on Shift 3 were subtracted from Shift 2. This permitted direct comparisons of transfer scores. Figure 1 is a plot of mean difference or transfer scores as a joint function of type and order of shifts. The summarized analyses of variance are presented in Table 3. Differences among the mean transfer

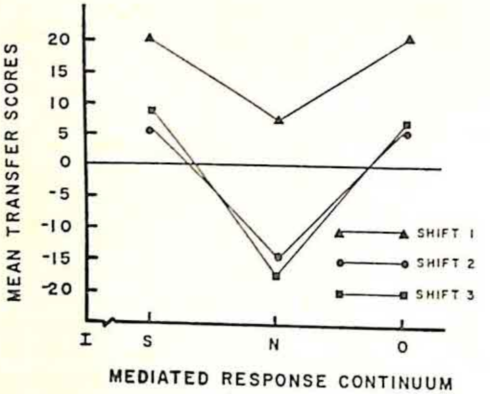


FIG. 1. Mean transfer scores (difference scores in successive trials to criterion) as a function of mediated response "similarity" with order of shifts as the parameter. (Transfer in concept attainment is shown after the preceding acquisition of concepts that were similar [S], neutral [N], or opposite [O]. Identical [I] represents the origin of the "similarity" classification.)

scores for types of shifts failed to achieve statistical significance on Shift 1; all transfer scores, however, were positive. Transfer scores for the three types of shift are statistically significant on Shifts 2 and 3, as are the quadratic components. Transfer is negative for Group N on Shifts 2 and 3. Although statistical comparison of Shifts 1, 2, and 3 are precluded by their lack of independence, inspection of Fig. 1 reveals a considerable negative displacement between the first and subsequent shifts. Results for the initial learning are essentially the same as those reported in Exp. I; absolute scores for the first shift are almost identical to the data reported in Tables 1 and 2.

TABLE 3

ANALYSIS OF VARIANCE OF TRANSFER SCORES FOR SHIFTS 1, 2, AND 3 OF EXP. II

Source of Variance	*df*	Shift 1 *F*	Shift 2 *F*	Shift 3 *F*
Shifts (S)	2	1.98	14.76***	28.01***
Linear	1	.00	.05	.27
Quadratic	1	3.95	29.46***	55.75***
Articulation (A)	1	.68	.95	.48
Instructions (I)	1	.00	.30	2.52
Stimulus Dimensions (D)	1	.14	.00	.08
S × A	2	2.32	.50	1.08
S × I	2	.05	7.60**	.87
S × D	2	3.30*	2.41	.48
A × I	1	2.07	1.86	.23
A × D	1	.05	.02	1.34
D × I	1	1.43	.95	2.36
S × A × I	2	.20	.85	1.15
S × A × D	2	.37	1.13	3.42*
S × I × D	2	1.39	.14	2.67
A × I × D	1	.89	2.86	2.09
S × A × I × D	2	.41	4.03*	.06
Residual (*MS*)	72	(374.65)	(281.11)	(246.08)

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Statistically significant interactions were obtained in each of the shifts; two of the interactions are plotted in Fig. 2.

Absence of statistical significance for S, N, O in Shift 1 is an artifact. Size concepts required significantly more trials for initial learning than brightness concepts (cf. results of Exp. I). The differences in difficulty level between the two relevant dimensions are subject to habituation and disappear after initial learning; a Shifts × Dimensions interaction is the result (Fig. 2, Shift 1). The basis for this interaction is documented in Table 4. Relevant dimensions are switched only in a Type N shift. Notice that the number of trials to criterion is higher for the more difficult size dimension during initial learning. This changes after the initial learning

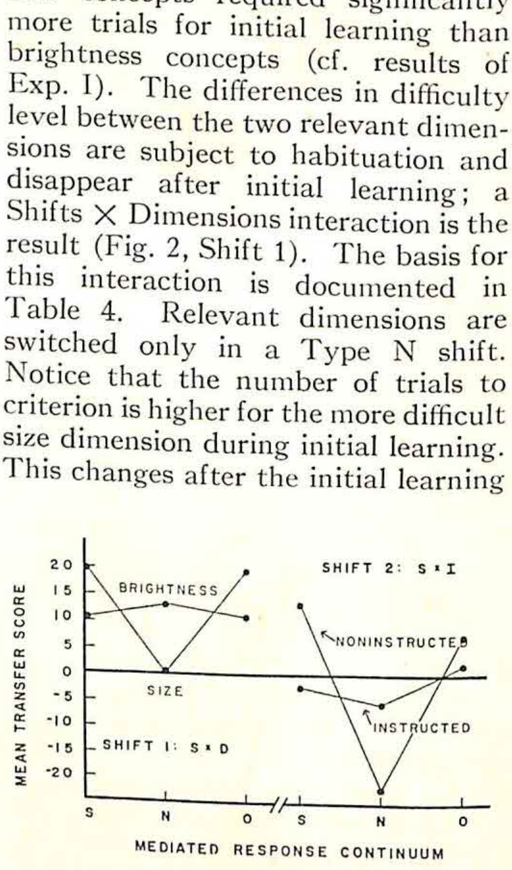


FIG. 2. Shifts × Stimulus Dimensions interaction for Shift 1 (left), and Shifts × Instructions interaction for Shift 2 (right).

TABLE 4

Explanation of the Shift × Stimulus Dimensions Interaction: Scores for the Neutral Shift (N)

Source	Stimulus Value	Mean	Computation
Initial learning: Trials to criterion	Size (S_1)	29.50	$\frac{\Sigma S_1}{n}$
	Brightness (B_1)	19.81	$\frac{\Sigma B_1}{n}$
Shift 1: Trials to criterion	Size (S_2)	19.19	$\frac{\Sigma S_2}{n}$
	Brightness (B_2)	16.00	$\frac{\Sigma B_2}{n}$
Transfer scores	Size	+.62	$\frac{\Sigma B_1 - \Sigma S_2}{n}$
	Brightness	+13.50	$\frac{\Sigma S_1 - \Sigma B_2}{n}$

and the difficulty level of the two relevant dimensions are roughly equal on the first shift (habituation effects). In addition, both dimensions are easier on the first shift following initial learning (task learning sets). Thus, subtracting the trials to criterion scores on Shift 1 from those for initial learning produces difference or transfer scores that are higher for brightness than for size in Cond. N of Shift 1 (compare Fig. 2, Shift 1 with Table 4). The result is a spurious interaction in Shift 1 that cancels the main effects for shifts.

Discussion

Concept shifts, such as the three types investigated, may be defined at several levels of discourse. Antecedent operations consist of switching the reinforcement from one to another stimulus value. The rate of change in *S*'s consequent choice behavior is recorded. A first-order mediational approach would assume that *S* covertly emits an appropriate verbal response in the presence of a given stimulus value. In actual fact, verbal mediation of the first order was something more than an assumption in both experiments. The *S*s were pretrained to emit verbal responses and half did so overtly throughout the study. A second-order mediational approach appeals to response classes and hierarchies, usually verbal. One mediational paradigm (Kendler & Kendler, 1962) implies that any intradimensional shift (e.g., S or O) should involve the same second-order mediator. If the postshift concepts are of equal difficulty or counterbalanced and the second-order mediator is the same, then the rates of postshift learning should be the same. Both absolute learning and transfer rates were virtually identical for concepts rated S and those rated O. Obtained transfer effects as a function of mediated response similarity, however, suggest two opposed explanatory hypotheses: direct word association and verbal associative commonality. Haagen (1949) reported high correlations between ratings of synonymity and ratings of associative connection. Cofer (1957) demonstrated increasing associative commonality (i.e., the number of common associations to each word in word pairs) with ratings of increased similarity. Present similarity ratings, therefore, imply a high associative connection between the S mediators. The O and S mediators may be functionally equivalent since antonyms are usually primary associates (Russell & Jenkins, 1954). The direct association hypothesis states that prepotent S-R connections produce the transfer effects in concept shifts. After learning a "black" concept, for example, the high probability of implicit emission of the word "white" is suggested by word association tests. This highly parsimonious hypothesis may account for a variety of data while assuming no more than first-order mediation. The relationship between age and reversal shift responding (Kendler & Kendler, 1962) may, in part, be due to characteristic age differences in word association. Children give considerably fewer opposite associations than do adults (Woodworth & Schlosberg, 1954). Transfer data for the S shifts indicate an apparent difficulty for this hypothesis: the frequency of gray as a response to white and black is, respectively, 1/1008 and 0/1008 in the Minnesota Norms

(Russell & Jenkins, 1954). Contextual restraints on word associations produced by the experimental arrangements may markedly alter association norms. If this cannot be demonstrated, then the associative commonality hypothesis would seem to offer a superior alternative for the present data. The associative commonality hypothesis states that transfer between concepts will covary with the degree to which mutual or common word associations occur to the two concept names. According to this hypothesis, the equal transfer obtained for S and O shifts results from equality in the number of mutual associates for concept names such as black and gray (S) and black and white (O). The present study does not permit a decision between the two hypotheses. The rating data, however, suggest general bidirectional relationships in similarity ratings and hence in associations. The frequent unidirectional relationships observed in word association tests may be a serious limitation of the direct association hypothesis.

The two hypotheses outlined may be equally germane to studies of response similarity and transfer or retroaction in verbal learning. Comparison of the present data to transfer studies employing verbal paired associates would be less difficult with a paired-associate concept learning design (Goss, 1961). This design would have added unnecessary complications to the concept attainment process. An analysis in terms of verbal learning data, nevertheless, may be instructive. The significant quadratic effect initially appears consistent with the findings of Bugelski and Cadwallader (1956); Osgood's (1946) study implies a linear trend. That is, the superior transfer in O shifts as compared with N agrees with the Bugelski and Cadwallader data but is inconsistent with Osgood's results. Category O in Osgood's study consisted of "opposed" words rather than antonyms, and very likely involved a low level of direct or common association. Bugelski and Cadwallader employed Osgood's opposed list but Underwood (1961) reports a methodological difficulty in their experiment. Accelerated learning in the present study for S shifts as compared with N is consistent with the general finding that transfer and retroactive facilitation increases with response similarity (Gladis & Braun, 1958; Osgood, 1946; Young, 1955).

Instructions concerning a shift in reinforcement produced no significant main effects and one highly significant interaction (Table 4, Shift 2). The instructed group immediately extinguishes (i.e., abandons) the incorrect response when they are informed of a shift. The near zero transfer effects obtained for the instructed group (Fig. 2, Shift 2) suggest that simple concept transfer may consist largely of the extinction of the old response. Repeated interactions of shifts and instructions would support this view; the reasons for the absence of a shifts and instructions interaction in Shift 3 are obscure.

Although the proposed relationships between sequential concept learning and verbal behavior cannot be viewed as definitive, the results are sufficiently encouraging to warrant additional effort. In essence, the axiom that generated the research reduces to this: if it is assumed that a mediator is verbal, then the laws governing mediation should approximate the empirical relationships of verbal behavior.

REFERENCES

BUGELSKI, B. R., & CADWALLADER, T. C. A reappraisal of the transfer and retroaction surface. *J. exp. Psychol.*, 1956, **52**, 360–366.

COFER, C. N. Associative commonality and rated similarity of certain words from Haagen's list. *Psychol. Rep.*, 1957, **3**, 603–606.

GLADIS, M., & BRAUN, H. W. Age differences in transfer and retroaction as a function of intertask response similarity. *J. exp. Psychol.*, 1958, **55**, 25–30.

GOSS, A. E. Verbal mediating responses and concept formation. *Psychol. Rev.*, 1961, **68**, 248–274.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of

400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychol. Rev.*, 1962, **69**, 1–16.

OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277–301.

OSGOOD, C. E. The similarity paradox in human learning. *Psychol. Rev.*, 1949, **56**, 132–143.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract No. N8onr-66216, Office of Naval Research.

UNDERWOOD, B. J. An evaluation of the Gibson theory of verbal learning. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

YOUNG, R. K. Retroactive and proactive effects under varying conditions of response similarity. *J. exp. Psychol.*, 1955, **50**, 113–119.

(Received January 12, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 30–34

HYPNOSIS IN THE UNHYPNOTIZABLE: A STUDY IN ROTE LEARNING[1]

DAVID ROSENHAN

Personality Research Group, Educational Testing Service, Princeton

AND PERRY LONDON[2]

University of Illinois

This study was concerned with the effects of hypnosis on the memorization of nonsense syllables. 32 undergraduates, 16 relatively susceptible and 16 relatively unsusceptible, were individually presented 10 nonsense syllables on a memory drum under counterbalanced hypnotized and unhypnotized conditions. Findings demonstrated that hypnosis facilitates the performance of the relatively unsusceptible Ss, and impairs the performance of the very susceptible ones. Hypnosis was thus seen as having potentially variable effects which interact with susceptibility and with the specific experimental measure which is used.

Both "common sense" and the bulk of the literature on hypnosis assume that hypnotic induction procedures affect very "susceptible" people alone. There is, therefore, a substantial literature dealing with who can and who cannot be hypnotized. Nevertheless, experimental evidence to the effect that hypnosis also influences the behavior of unsusceptible *S*s has been available for more than a generation.

In studying the effects of hypnosis on the memorization of meaningless materials, Young (1925) discovered that control *S*s who, though relatively unsusceptible to hypnosis, were subjected to the induction procedure, performed distinctly better as a group than did susceptible *S*s in the hypnotic condition.

This interesting finding has been generally neglected, perhaps because it appears to contradict conventional beliefs about hypnosis. Young, himself, in reporting his experiment, dismisses the effects of his induction procedure on controls as "pseudohypnosis," and Weitzenhoffer (1953), reviewing the literature, proposes without explanation that these effects "probably have nothing to do with hypnosis" (p. 171).

Such a dismissal is justified if one views hypnosis as a mystical or metaphysical state which is discretely available to some people but not to others. If, however, one takes the position that hypnosis can only be defined in terms of specific responses to criterion procedures which we agree to call hypnosis, then Young's study (which employed only 3 control *S*s) requires replication.

Other studies of the effects of hypnosis on learning, in addition to disregarding Young's early work, have had several limitations (Eysenck, 1941; Sears, 1955; White, Fox, & Harris, 1940). As London and Fuhrer (1961) note, the degree of *S*s' hypnotic susceptibility was not determined in most studies. Also, *S*s often were not examined in both hypnotic and waking states for control purposes. Finally, since experimental *S*s were exposed to the hypnotic procedure while control *S*s were not, one cannot be certain that different group performances were due to the effects of the hypnotic manipulation rather than to the fact that one group got the subtle equivalent of a pep talk while the other did not. London and Fuhrer report an experimental design which permits the study of the interaction of hypnosis and susceptibility while avoiding some of the aforementioned pitfalls.

---

[1] The study was supported in part by a grant from the Haverford College Faculty Research Committee and by an NSF Research Fellowship to the first author. The authors are indebted to Gary Holtzman of Haverford College and Leslie M. Cooper of the University of Illinois, who, respectively, assisted in the conduct of the study and analysis of data. Gerald S. Blum and Francis W. Irwin made many helpful criticisms of theory and manuscript.

[2] Currently on leave, Department of Psychology, Stanford University.

The present study applied this design to the study of the effects of hypnosis on rote learning, replicating Young's experiment thereby.

METHOD

Subjects

The *S*s were 32 male undergraduates, mostly freshmen and sophomores 17–20 yr. old, who were selected from among 68 volunteers in an introductory psychology course.

All volunteers were administered an unpublished test known as the Group Hypnotic Susceptibility Scale (GHSS). The items and procedures of this scale are derived from Forms A and B of the Stanford Hypnotic Susceptibility Scale (Weitzenhoffer & Hilgard, 1959), with appropriate conservative alterations which permit its use with groups of up to 5 *S*s. Both the GHSS and the Stanford scale are standardized hypnotic induction procedures. These procedures include the familiar exhortations to relax and 12 tasks (such as postural sway, arm rigidity, hallucination, amnesia, etc.) which the *S* is asked to perform. Degree of susceptibility to hypnosis is measured by the number of items *S* performs.

Two categories of *S*s were selected for the experiment by means of the scale: Tranceable (T), i.e., highly susceptible to hypnosis, and Untranceable (UT), i.e., relatively unsusceptible to hypnosis. In order for *S* to qualify in the T category, it was necessary for him to pass 8 or more of the 12 items on the scale, including at least one of the four most difficult items (arm immobilization, verbal inhibition, eye catalepsy, amnesia). To qualify in the UT category, *S* had to fail 8 or more of the 12 items, including all of the most difficult ones. Sixteen T *S*s and 16 UT *S*s selected by these means actually completed the experiment.

Apparatus

Two lists of 10 consonant-vowel-consonant nonsense syllables of 53% association value (Glaze, 1928) were presented to *S*s on a Gerbrands memory drum, one list for each condition. Each trigram on the list was exposed to view for 1 sec., with 2 sec. between exposures. Each list was presented three times, with a 5-sec. interval between presentations.

Procedure

Screening session.—After the administration of the GHSS, *S*s who failed to meet criteria were discharged. All *S*s were asked not to discuss the procedures with anyone until the experiment was complete.

Those *S*s whose GHSS performance placed them in either the T or UT category were rehypnotized. The instructions which *E* read for this second induction were taken essentially from Form B of the Stanford scale. Among the remarks contained in this procedure were statements to the effect (*a*) that for the purpose of this experiment, the *S*s were good hypnotic subjects, (*b*) that they would be hypnotized again in later sessions, and (*c*) that hypnosis would, in the future, be induced by a simple signal rather than a long and involved procedure. The verbal signal then described was used verbatim for the later induction of hypnosis during the experiment. After arousing the *S*s, *E* arranged individual sessions for each *S*. These took place from 2 to 8 days after the screening session.

Experimental session.—The second session constituted the experiment proper. It consisted of three parts: (*a*) Base rate learning of nonsense syllables, (*b*) hypnotic learning of nonsense syllables, and (*c*) administration of the Stanford scale, Form B. Conditions *a* and *b* and Syllable Lists *a* and *b* were counterbalanced. In every case, *S* took several physical performance tests prior to the learning task. In the sequence of experimental conditions, the presentation of learning tasks occurred 15 min. apart. Because *S*s served as their own controls in this counterbalanced experiment, and in order to hold nearly constant the amounts of time *S*s spent in hypnosis, the fixed number of presentations method was preferred to learning to criterion.

The *S*s were told that they would see the entire list three times, after which they would be asked to spell as many of the syllables as they could remember. No further comments were made to *S*s; specifically, suggestions to the effect that hypnosis would facilitate their learning were avoided, in order to gain information about the effects of hypnosis per se upon memorization.

Upon completion of the experimental tasks, Form B of the Stanford Hypnotic Susceptibility Scale (SHSS) was administered to see if *S*'s susceptibility score had shifted during the course of the experiment. Subsequently, *S* was told that the experiment was concluded, asked to report his subjective reactions, and then discharged.

RESULTS

The *S* was assigned a memorization score for each experimental condition which consisted of the total number of trigrams which he spelled correctly after three presentations of the entire list. Mean scores of the two groups for the different conditions are presented in Table 1.

The data were subjected to analysis of variance (Lindquist, 1953, p. 269). While the T-UT dimension was not significant, the Hypnotized-Unhypnotized conditions differed considerably from each other ($p < .01$). There was, moreover, a marked interaction between dimensions ($p < .001$) such that the performance of T *S*s declined during hypnosis while that of UT *S*s improved during hypnosis (Table 2). Tests of simple effects showed that the UT group remembered significantly more syllables when hypnotized than when not hypnotized ($p < .05$), but the decline in performance of the T group from the Unhypnotized to Hypnotized conditions was not significant at the .05 level. Further inspection of the data revealed that for UT *S*s, hypnotic memorization of nonsense syllables was always equal or superior to base rate performance; there were no decrements. For T *S*s, hypnotic increments were found in only three cases, while decrements during the hypnotic condition occurred in eight cases.

TABLE 1

MEANS AND *SD*s FOR MEMORIZATION OF NONSENSE SYLLABLES

Condition	Group T		Group UT	
	Mean	*SD*	Mean	*SD*
Base Rate	7.38	1.82	5.62	1.20
Hypnotized	6.88	1.54	7.50	1.55

TABLE 2

ANALYSIS OF VARIANCE FOR NONSENSE SYLLABLES MEMORY SCORES

Source	*df*	*F*
Between-*S*s	(31)	
Tranceable vs. untranceable (T)	1	1.01
Error (between-*S*s *MS*)	30	(4.05)
Within-*S*s	(32)	
Hypnotized vs. unhypnotized (H)	1	10.36**
H × T	1	30.92***
Error (within-*S*s *MS*)	30	(0.73)

** $p < .01$.
*** $p < .001$.

Scores on Form B of the Stanford scale ranged from 1 to 7 (mdn. = 3.50) for UT *S*s, and from 6 to 12 (mdn. = 9.25) for T. In no instance did an *S* designated UT on the GHSS become T on the Stanford scale, or vice versa. Since the score differences between the tests are in the same direction and of about the same magnitude as differences between administrations of Forms A and B of the Stanford scale itself (London & Fuhrer, 1961) under similar circumstances, it seems reasonable to infer that screening data obtained with the GHSS is roughly equivalent to that obtained with Form A of the Stanford scale.

DISCUSSION

The results of this study are manifestly surprising. They suggest, first of all, that there are some initial differences in the learning ability of the very susceptible and the relatively unsusceptible. But more important, they demonstrate that hypnosis facilitates the performance of the relatively unsusceptible *S*s, and impairs the performance of the very susceptible ones.

The finding of initial differences in learning ability between T and UT *S*s is reminiscent of previous studies in which base rate differences on physical performance tasks were reported (London & Fuhrer, 1961; Rosenhan & London, 1963). In these studies, however, differences consistently favored UTs, and led to the hypothesis of possible subtle personality differences between T and UT *S*s, an hypothesis which is reaffirmed here. The matter is, however, by no means clear; the fact that base rate differences in this study favor Ts suggests that the effects of hypnosis are not only *S* specific, but also task specific.

That hypnosis should facilitate the learning of UTs and relatively impair the performance of Ts is difficult to explain. One possible hypothesis is that there is an interaction between the characteristic relaxation which people experience when hypnosis is induced, some hypothetical optimum level of tension which is required to memorize effectively, and a possible personality difference between T and UT *S*s. If so, one might argue as follows: UT *S*s are, to begin with, *too tense* to learn very well, but T *S*s are only *optimally tense* and therefore give their best performances. When hypnosis is induced, however, T *S*s relax to the point where they become inattentive, while UT *S*s relax to the point where they become more appropriately attentive. If anxiety or drive facilitates performance up to a certain intensity and impairs it beyond that intensity, the argument has some plausibility.

But this interpretation is clearly speculative, and the importance of these results does not lie in the temptation they offer for intellectual virtuosity, but in their implications for research on hypnotic phenomena. It is now clear that the assumption that an untranceable *S* is not affected by the hypnotic induction is no longer tenable. The estimate of susceptibility which emerges from a scale of hypnosis (such as the GHSS or SHSS) is one of limited parameter; it does not necessarily predict *S*'s performance on nonscale tasks.

One implication of these findings is that the conclusion of London and Fuhrer (1961) that "the effects of hypnosis cannot be very effectively studied as a quantum independent of the relative susceptibility of *S*s" now appears much too limited. The data presented here indicate that hypnosis is a situation with potentially variable effects which interact per susceptibility status *per measuring instrument*. Thus, it is not surprising that different experiments in the same area produce conflicting results both since previous research has not commonly concerned itself with the susceptibility dimension and since generalizations about the effects of hypnosis have been offered without much consideration of the limitations of different measuring instruments. This argument suggests that it is improper to define hypnosis for experimental purposes or to generalize about its effects without consideration of the susceptibility of the *S*s and the limited scope of the experimental tasks.

REFERENCES

EYSENCK, H. J. An experimental study of the improvement of mental and physical functions in the hypnotic state. *Brit. J. med. Psychol.*, 1941, **18**, 304–316.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

LONDON, P., & FUHRER, M. Hypnosis, motivation and performance. *J. Pers.*, 1961, **29**, 321–333.

ROSENHAN, D., & LONDON, P. Hypnosis: Expectation, susceptibility, and performance. *J. abnorm. soc. Psychol.*, 1963, **66**, 77–81.

SEARS, A. B. A comparison of hypnotic and waking learning of the International Morse Code. *J. clin. exp. Hypn.*, 1955, **3**, 215–221.

WEITZENHOFFER, A. M. *Hypnotism: An objective study in suggestibility.* New York: Wiley, 1953.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale.* Palo Alto: Consulting Psychologists Press, 1959.

WHITE, R. W., FOX, G. F., & HARRIS, W. W. Hypnotic hypermnesia for recently learned material. *J. abnorm. soc. Psychol.*, 1940, **35**, 88–103.

YOUNG, C. An experimental study of mental and physical functions in the normal and hypnotic state. *Amer. J. Psychol.*, 1925, **36**, 214–232.

(Received January 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 35–38

DIFFERENTIAL EYELID CONDITIONING UNDER EQUATED DRIVE AS A FUNCTION OF THE REINFORCING UCS[1]

KENNETH W. SPENCE AND BLAINE F. TANDLER

State University of Iowa

The effects of varying the reinforcing properties of the paired UCS on differential conditioning under equated drive level (D) was studied. The high reinforcement (HR) group had the positive CS (500-cps tone) paired with a strong UCS (2.0 psi) at an optimal CS-UCS interval (500 msec.) while the negative CS (5000-cps tone) was paired with a weak UCS (.33 psi) at interval of 2550 msec. These conditions were reversed for low reinforcement group (LR). HR group performed significantly higher than LR group both for positive and negative CS. Differentiation of the discriminanda was greater for HR group. Since drive levels of 2 groups were equated the results may be interpreted as indicating greater amount of habit strength for both stimuli in case of HR.

A number of recent studies have been concerned with the question of whether in classical aversive conditioning the associative or habit factor (*H*) is a function of the intensity of the reinforcing UCS (Spence, 1953; Spence, Haggard, & Ross, 1958; Trapold & Spence, 1960). Employing an experimental design that used two levels of the UCS for groups of *S*s, one group (high reinforcement) had a strong UCS paired on conditioning (i.e., *H* producing) trials with the CS, while on the other half of the trials (nonconditioning) a weak UCS was presented alone, or as in one experiment, the weak UCS was paired with the CS at a CS-UCS interval (2.5 sec.) known not to produce eyelid conditioning in humans. In the case of the low-reinforcement group these conditions were reversed, i.e., the CS was paired on conditioning trials with the weak UCS, while on nonconditioning trials the strong UCS was used. Since the drive levels (*D*) of the high- and low-reinforcement groups were equated, the significantly higher performance level of the high-reinforcement group in each experiment may be interpreted as reflecting the development of a greater amount of the learning factor (*H*) in the high-reinforcement condition than in the low-reinforcement condition. This evidence of the development of differential *H* was taken as supporting a reinforcement type theory that habit strength (*H*) in such classical aversive conditioning is a function of the UCS.

In the present study this theoretical interpretation was tested further by applying the experimental procedure of equating drive level, while at the same time varying the reinforcing property of the UCS, to the differential conditioning situation. Thus, in the case of the high-reinforcement group the positive CS was always paired at an optimal CS-UCS interval with a strong UCS, while the negative CS was followed at a long, nonconditioning interval with a weak UCS. For the low-reinforcement group the two different UCSs were reversed.

---

[1] This research was carried out as a part of a project concerned with the influence of motivation on performance in learning under Contract N9 onr-93802, Project NR-154-107 between the State University of Iowa and the Office of Naval Research. Acknowledgment is made to Paul Butler and Edward Martin, who ran some of the *S*s.

Under these conditions our theory would lead to the expectation that the high-reinforcement group would give a higher percentage of CRs to the positive CS than would the low-reinforcement group. The implications of the theory for the negative CS and for the degree of differentiation are more complicated and will be treated in the later discussion section.

METHOD

Subjects.—Eighty women from an introductory course in psychology served as *S*s. An additional 14 *S*s were discarded for various reasons. Five were eliminated for giving CRs on pretest presentations of the tones, 2 for failure to give a single CR to the positive stimulus during conditioning, and 7 were voluntary responders (Spence & Ross, 1959).

Apparatus.—The equipment for recording eyelid closure and presenting the UCS was the same as that used in previous studies from this laboratory (cf. Spence & Taylor, 1951). The CS was a tone from a loudspeaker driven by one of two Hewlitt-Packard oscillators, as selected by *E*. A 500-cycle tone served as the positive CS (S^+) and a 5000-cycle tone as the negative CS (S^-). In the area of *S*'s head, both tones had a sound pressure level of 70 db. as measured by a General Radio sound level meter, Type 759-B. Tone durations were 550 msec. for S^+ and 2550 msec. for S^-. The air puffs were of 50-msec. duration and occurred during the last 50 msec. of either CS interval, terminating with the CS. The longer CS-UCS interval has been shown to produce little or no conditioning (McAllister, 1953a, 1953b) and, when employed in the differential conditioning situation with S^-, to lead to differential level of response just as omission of the UCS does (Runquist, Spence, & Stubbs, 1958).

Two UCS intensities, a strong puff of 2 psi and a weak puff of .33 psi, were used in the conditioning of each *S*, with S^+ always followed by one puff strength and S^- by the other.

A response deflecting the recording pen 1 mm. or more during the interval 150 to 500 msec. following the onset of either S^+ or S^- was scored as a CR. Responses with a latency less than 150 msec. were scored as original responses to the tone and were not included in the data.

Procedure.—On each trial, a verbal ready signal was given 2–4 sec. before onset of the CS. All *S*s were instructed to blink once upon presentation of the ready signal and then to fixate the disc until the tone terminated. Five pretest trials, consisting of the presentation of S^+, S^+, S^-, S^- and the UCS alone, were given. Groups receiving two puff strengths had the strong UCS alone. Immediately following pretest, the conditioning period was begun. Both S^+ and S^- were presented five times within each block of 10 trials, according to a prearranged schedule, with the restriction that no more than 3 S^+ nor more than 2 S^- trials occurred in succession. Intertrial intervals were either 15, 20, or 25 sec., by fixed schedule, with an average of 20 sec.

Experimental design.—Each *S* was assigned to one of two groups in order of appearance in the laboratory. For the *S*s of Group S (strong reinforcement) S^+ was always followed by the strong puff and S^- by the weak puff. This combination was reversed for Group W (weak reinforcement) with S^+ followed by the weak puff and S^- by the strong puff.

RESULTS AND DISCUSSION

Performance to the positive CS.—Acquisition curves of conditioning to both positive and negative stimuli are presented in Fig. 1 in terms of the

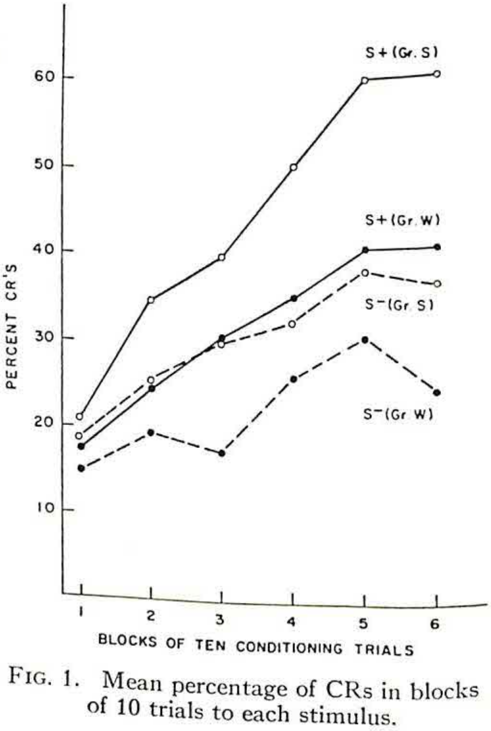


FIG. 1. Mean percentage of CRs in blocks of 10 trials to each stimulus.

percentage of CRs in each block of 20 trials, 10 of which were S^+ trials and 10 S^- trials. As is evident, the performance of Group S was higher than that of Group W on both S^+ and S^- trials. Evaluation of the performance levels to S^+ over all trials by the Mann-Whitney U test indicated that the difference between the two groups was significant at the .01 level for a one-tailed test. Over the last 20 trials the difference by the same test was found to be significant at the .001 level. This finding confirms the prediction made from the reinforcement type theory described in the introduction. According to this interpretation the higher performance of Group S implies that a higher level of H is developed to the positive S in this group than in Group W as the drive levels of the two groups were equalized.

Performance to the negative CS.—On the basis of what is known about generalization of conditioning it may be inferred that the strength of the generalized habit strength ($\bar{H}$) to S^- would be greater for Group S than for Group W. The implication of this, neglecting all other factors, is that Group S should also exhibit a higher level of performance to the negative CS than Group W. As Fig. 1 shows, this expectation was confirmed. Although the difference is somewhat less than in the case of the positive CS, it is significant at the .04 level (Mann-Whitney U test, two-tailed) for all 60 trials and also for the last 20 trials.

Nevertheless, a complicating factor must be considered in the case of the negative CS, namely the magnitude of the decremental, inhibitory factor (I) resulting from nonreinforcement of this CS. Unfortunately, we have little knowledge concerning the variables that determine I in classical aversive conditioning. If I is a positive function of E, the effective excitatory strength ($\bar{E}$) of the response to the negative CS would be weakened more in the case of Group S than Group W. Such an effect, it should be noted, would account for the obtained finding that the difference in performance between Groups S and W was less in the case of the negative CS than for the positive CS.

Differentiation of response strength to S^+ and S^-.—According to our theory, the degree of differentiation as reflected by the difference in the response strengths to the positive and negative stimuli is directly related to the difference between their excitatory potentials. The latter may be shown to be as follows for Groups S and W:

$$\begin{aligned} E_s{}^+ - \bar{E}_s{}^- \\ &= D(H_s - \bar{H}_s) + I_s(\text{Group S}) \end{aligned}$$

$$\begin{aligned} E_w{}^+ - \bar{E}_w{}^- \\ &= D(H_w - \bar{H}_w) + I_w(\text{Group W}) \end{aligned}$$

As may be seen, since the value of D is the same for both groups, the degree of differentiation will depend, theoretically, on the magnitude of the differences in the values of I_s and I_w and the values in the parentheses. The latter represent the decrements in habit strengths due to generalization. The finding of this experiment that differentiation was significantly greater in Group S than Group W (at the .03 level by the Mann-Whitney U test, two-tailed) implies that either one or both of these factors are larger in the case of Group S than Group W. The present data do not permit any inference as to which of these two factors might be responsible for the difference in differentiation or whether both are involved.

It is interesting to note that if a strong, 2-psi puff had been used for *both* discriminanda in one group and a weak, .33-psi puff for *both* stimuli in the other, the equations for $E^+ - \bar{E}^-$ would differ from those given above in that the values of D in the two equations would differ, i.e., $D_s > D_w$. It follows from these two sets of equations that the *difference* in degree of differentiation $E_s^+ - \bar{E}_s^- - (E_w^+ - \bar{E}_w^-)$ should be greater for groups run under this condition than for groups run under the conditions of the present experiment. Some evidence against this implication of the theory is provided by a comparison of the results of the present study with those obtained in an earlier investigation of differential conditioning (Runquist et al., 1958). In the latter study a 2-psi puff was used for both S^+ and S^- in one group (S) and a .33-psi puff for both stimuli in the other group (W). Otherwise, the conditions of the two studies were as similar as they could be made. The degree of differentiation in terms of the difference in the mean percentage of responses made in the last 20 trials to the positive and negative stimuli ($S^+ - S^-$) is shown in Table 1. As is apparent, the implications of the theoretical equations are only partially supported. While, as the theory demands, the degree of differentiation was greater for the S groups than the W groups in both experiments, the implication that difference between the differentiations would be greater for the 1958 study than for the 1960 study was not confirmed.

TABLE 1

COMPARISON OF RESULTS OF PRESENT STUDY WITH THOSE OF RUNQUIST, SPENCE, AND STUBBS (1958)

Groups	Degree of Differentiation ($\%S^+ - \%S^-$)	
	Present Study	1958 Study
S	23.1	25.6
W	12.1	21.6
Difference (S−W)	11.0	4.0

The major discrepancy in the data is the relatively high differentiation shown by Group W of the 1958 study. The level of conditioning of this group to the positive CS was much higher than that usually obtained with this level of puff intensity. If one uses only the middle 50% of the *S*s in all four groups, the differences between the differentiation were more in line with the theory. The differences in this case were 13.8 for the 1958 study and 9.2 for the present one.

REFERENCES

McALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-US interval. *J. exp. Psychol.*, 1953, **45**, 423–428. (a)

McALLISTER, W. R. Eyelid conditioning as a function of the CS-US interval. *J. exp. Psychol.*, 1953, **45**, 417–422. (b)

RUNQUIST, W. N., SPENCE, K. W., & STUBBS, D. W. Differential conditioning and intensity of the UCS. *J. exp. Psychol.*, 1958, **55**, 613–616.

SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

TRAPOLD, M. A., & SPENCE, K. W. Performance change in eyelid conditioning as related to the motivational and reinforcing properties of the UCS. *J. exp. Psychol.*, 1960, **59**, 209–213.

(Received January 19, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 39–46

INFORMATION PROCESSING UNDER CONTRADICTORY INSTRUCTIONAL SETS[1]

WILLIAM C. HOWELL AND DAVID L. KREIDLER

Ohio State University

Key pressing in response to spatially coded visual signals was studied as a function of 4 sets of instructions: (a) speed-accuracy, (b) speed, (c) accuracy, and (d) transmission rate (H_t/sec). Contradictory (speed-accuracy) and noncontradictory (H_t/sec) instructions resulted in comparable acquisition and performance functions on all 3 measures taken: speed, accuracy, H_t/sec. Behavior under conflicting instructions corresponded more closely to that for accuracy than to that for speed; S appeared to select a high accuracy criterion and to adjust speed to the maximum level permitted by it. Speed instructions caused S to lower his accuracy criterion appreciably, but with little compensation in increased speed.

Edwards (1961) has pointed out that instructions used in many psychological experiments are ambiguous or internally contradictory. Frequently, *S* is instructed to maximize two or more quantities each of which requires a different, and generally incompatible, way of responding. Particularly susceptible to this criticism are investigations of human information-processing capabilities. The instructions commonly used emphasize both speed and accuracy of responding (e.g., Alluisi & Muller, 1958; Hick, 1952; Howell, 1962; Klemmer, 1958) which goals are, of course, directly antagonistic. Knowledge of results, which serves as a supplement to instructions, frequently is not given or, even worse, is misleading. This is particularly true of studies in which rate of information transmission (H_t/sec) is the primary measure of performance, for *S* is never informed of the fact that he is to be evaluated on this metric. As a result, he is unaware of the existence of a rule by which he should weight speed and accuracy, and he cannot realize that so long as he responds consistently he need not even be accurate.[2]

Faced with conflicting or ambiguous instructions, each *S* must assign his own weights to the alternative goals. To the extent that *S*s are alike in their weightings, instructional sets could be confounded with experimental treatments; to the extent that they are different, error estimates are inflated and real effects could be obscured. Thus, it is apparent that contradictory instructions could have an important bearing upon the outcome of information-processing experiments.

One solution to the problem of contradictory instructions has been proposed by Edwards (1961): this is to specify for *S* ". . . exchange rates among value dimensions which reduce them to a single final value measure" (p. 283). Money, he suggests, is the most useful such final value measure by virtue of the fact that it is used and understood by all potential *S*s. Whereas money is undoubtedly a common evaluative yardstick in our culture, it may also be a somewhat elastic one (Ashley, Harper, & Runyon, 1951), and hence rather undesirable for use in some psychological experiments. Another measure which is particularly worthy of consideration for information-processing studies is that of H_t/sec itself. This is, after all, the single quantity upon which performance is ultimately evaluated. Some investigators (e.g., Alluisi & Martin, 1958; Garvey, 1957) have avoided conflicting instructions by stressing speed or accuracy alone while continuing to use H_t/sec as the measure of performance. Such instructions, while unambiguous, would be expected to elicit specific and divergent modes of behavior.

---

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported by the United States Air Force under Contract No. AF 33(616)-6166, monitored by the Aerospace Medical Laboratory. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] The information measure H_t/sec is sensitive to speed and consistency, but not necessarily to accuracy of responses (Attneave, 1959).

The argument for a single common value to which all conflicting goals are reduced appears to be a sound one from a logical standpoint; it is important to recognize, however, that there is little empirical evidence to indicate the sensitivity of information-processing performance to contradictory instructions. It was the purpose of the present study, therefore, to determine precisely how contradictory instructions influence performance on a typical information-processing task. More specifically, a simple key pressing task was investigated under four sets of instructions: one set placed equal emphasis upon response speed and accuracy (i.e., a conflicting situation), a second set emphasized speed only, a third accuracy only, and a fourth H_t/sec. The latter three sets of instructions were considered to be noncontradictory in that *S* was required to maximize a single quantity. By comparing various characteristics of performance under these conditions, it was hoped that at least two questions might be answered: (*a*) Does performance under contradictory speed-accuracy instructions differ from that for comparable noncontradictory (H_t/sec) instructions? (*b*) Does performance under contradictory speed-accuracy (and, in fact, noncontradictory H_t/sec) instructions more closely approximate that for speed alone or that for accuracy alone? The former question asks, in a sense, whether or not one may accept the information measure as an index of performance obtained under speed-accuracy instructions. The latter question seeks to locate both instructional sets upon a continuum defined by pure speed emphasis at one end and pure accuracy emphasis at the other.

METHOD

Subjects.—Forty male students from the introductory psychology program at Ohio State University served as *S*s; they were divided randomly into four treatment groups of 10 *S*s each.

Apparatus.—The stimulus ensemble consisted of 10 neon glow lamps positioned directly above a set of 10 response keys. Coding was in terms of spatial contiguity: the appropriate response to the occurrence of any light was the depression of that key located directly beneath it.

The basic unit for presenting stimuli and recording responses was a punched-tape typewriter system (the Friden Flexowriter Programatic). As conventionally used, this system "reads" a code punched into tape and generates pulses which activate the appropriate characters on the keyboard. The system also contains a punch which permits the preparation of tapes by manual operation of the keyboard. For the present study, this basic system was modified in two ways: (*a*) the pulses generated by each signal on the tape were fed into an additional (external) reading circuit which selected the appropriate thyratron tube for the presentation of a particular stimulus light, and (*b*) *S*'s response keys were wired to the punch mechanism so that all responses could be preserved on tape.

Several additional features of the apparatus were concerned with pacing and response timing. The utilization of a "stop" code after each signal on the tape made it possible for the task to be self-pacing. Coincident with the presentation of each stimulus, a Standard Electric 1/100-min. timer was activated which continued to run until a response was initiated. In this way a cumulative record of response time was kept for an entire stimulus sequence. The initiation of the stimulus sequence was also under direct control of *S*: a green light on the stimulus panel signaled the readiness of the apparatus, after which time *S* could initiate the first stimulus by pressing any key. A red light signaled the end of a sequence.

All programing and scoring apparatus was located in an experimental booth separate from that in which *S* was seated. The *S*'s booth was small, painted white, and illuminated dimly by diffuse light from an outside source. The display panel and response keys were attached to a 30-in. high table before which *S* was seated; the viewing distance was approximately 28 in.

Procedure.—Each of four groups performed the same task under a different set of instructions. The task required *S* to press keys in response to the occurrence of single-light stimuli according to one of four sets of rules. Since instructions, complemented by summary feedback, thus constituted the major variable of the study, those portions specific to each group are quoted below. For all groups the first paragraph was an identical description of the general task; the second paragraph began, "You are to respond . . ." and ended in one of the following directions:

> [Group I] . . . as fast as possible; . . . although you would like to make correct responses, speed is the important thing and accuracy is definitely a secondary consideration.
>
> [Group II] . . . so as to make as few errors as possible; . . . speed is definitely a secondary consideration.
>
> [Group III] . . . as fast and as accurately as you can; . . . speed and accuracy are of equal importance in performing your task.
>
> [Group IV] . . . so as to convey information at the fastest possible rate. There is a way in which you can evaluate your information-conveying rate (i.e., come out with a score) which you must understand to do your best. We have 10 events (lights), any one of which could occur on any presentation. If you always press the same key when, say, this one comes on, you are conveying all the information available at that time. We can express this as a quantity: the *bit* is the unit used; one event selected from 10 can convey as much as 3.32 bits of information. The more you scatter your responses around, however, the less information you convey: if, for example, you used a *different* key each time this light came on, you would convey no information (zero bits). Similarly, if you pressed the *same* key no matter which light came on you would also convey no information (zero bits). What we are measuring, then, is the extent to which we can tell from your response which light came on: if you are consistent in your responses to the lights, a lot of information is conveyed (up to 3.32 bits per response); if you are inconsistent or use fewer responses than there are lights, the amount conveyed can decrease to zero bits. You can get a high score per response, then, by using one and only one key for each light—it should be easiest if you use the key directly below each one.
>
> There is another factor which enters into your score, and that is time. The faster you respond, the more bits you can get per time period for any level of light-key consistency. Your score, then, will be in *bits/sec.* Is this perfectly clear?

None of the *S*s appeared to encounter any great difficulty in grasping these essential concepts. It was recognized, of course, that *S*s in Group IV might interpret their instructions much as they would those for Group III (i.e., go as fast and as accurately as you can). Even so, however, the basic distinction between the two treatments remained: the unitary criterion of Group IV permitted *S* to choose and, if necessary, to modify, his approach on the basis of a single evaluative dimension; the conflicting criteria of Group III required, in addition, that *S* establish a "trade-off" between two dimensions.

The stimuli were presented in randomized sequences of 100 events each at a rate controlled by *S*. Each sequence constituted a trial, and 5 trials were presented successively in a block. All 20 trials were completed in one four-block session. Rest periods of approximately 1 min. occurred between trials, while periods of approximately 10 min. each separated blocks of trials.

Three response measures were recorded for all four groups: (*a*) number of errors per trial, (*b*) cumulative response time per trial, (*c*) mean H_t/sec per trial. Errors were determined by comparing *S*'s response tape with the corresponding input tape (it was necessary to print out both, one beneath the other, and to compare individual characters visually). Cumulative response time was read directly from the timer at the end of each trial. The H_t/sec scores were derived in the usual manner from an S-R matrix by subtracting equivocation entropy from the sum of the S and R entropies. The resulting H_t values were then divided by cumulative response times to yield H_t/sec scores (Attneave, 1959).

Summary knowledge of results was administered selectively to conform with the specific instructions: i.e., Group I received cumulative response time information; Group II received error information; Group III received both time and error information; and Group IV received H_t/sec information. Since the time required to obtain error and H_t/sec scores was rather great (approximately 2 min. per trial), knowledge of results was only provided on the last trial in each block; therefore, it reflected performance on only 400 of each 2000 stimulus presentations. Such a program of summary feedback appeared justified since its purpose was to maintain the original set rather than to strengthen specific associations. To facilitate the computation of H_t/sec feedback scores, a short-cut procedure was adopted in which estimates of H_t/sec could be read directly from a previously composed table. The assumptions required for this short cut were (*a*) that the most frequent response to each stimulus is the "correct" one, and (*b*) that all "errors" are randomly distributed. On any trial, then, the H_t/sec feedback score was determined by counting the number of errors, reading the appropriate H_t value from the table, and dividing by the cumulative response time. Since all errors were tabulated on an S-R matrix, it was possible to determine by inspection whether or not the assumptions of the short-cut procedure were tenable. In no instance was it deemed necessary to reject these assumptions, and later computation of H_t/sec scores by the conventional method revealed no serious discrepancies (never more than .01 bit/sec).

RESULTS

Performance was evaluated on the basis of the last block of trials only; the first three blocks (15 trials) were considered acquisition trials. The complete learning functions for all four groups are shown in Fig. 1. From these data it appears that all groups have achieved a reasonably slow rate of improvement by the fourth (performance) block of trials,[3] although it is probable that some improvement, particularly in response speed, would have continued for a number of trials thereafter (Mowbray & Rhodes, 1959).

Figure 1a indicates that there was little difference between any of the groups with respect to speed of responding; this conclusion was supported by the results of an analysis of variance applied to the performance data (group differences were nonsignificant). These data do suggest, however, that *S*s attempted to follow instructions, for the Group I (speed instructions) function is considerably higher than the Group II (accuracy instructions) function at all points.

The accuracy data summarized in Fig. 1b are of considerable interest in that they reveal no significant trend of improvement for any group over the complete 20 trials. Overall performance differences between groups, however, were large and statistically reliable ($p < .001$). The statistic used to test these differences was the nonparametric Kruskal-Wallis *H* test; its use was warranted by a highly significant heterogeneity of group variances as determined by Bartlett's test, $\chi^2(3) = 16.73$. This heterogeneity is attributable primarily to large individual differences occurring within Group I (see Table 1). Nonparametric two-sample tests (Mann-Whitney *U*) were carried out on the performance data to determine which pairs of groups differed in accuracy.

---

[3] The functions of Fig. 1 were plotted by blocks rather than by trials for the sake of clarity of inspection. This treatment of the data fails to express adequately the rapid improvement within Block 1 and the slow rate of improvement within Block 4 of Fig. 1a. Within Block 1 the linear slope of all curves was about .06; within Block 4 it did not exceed .004 for any group.

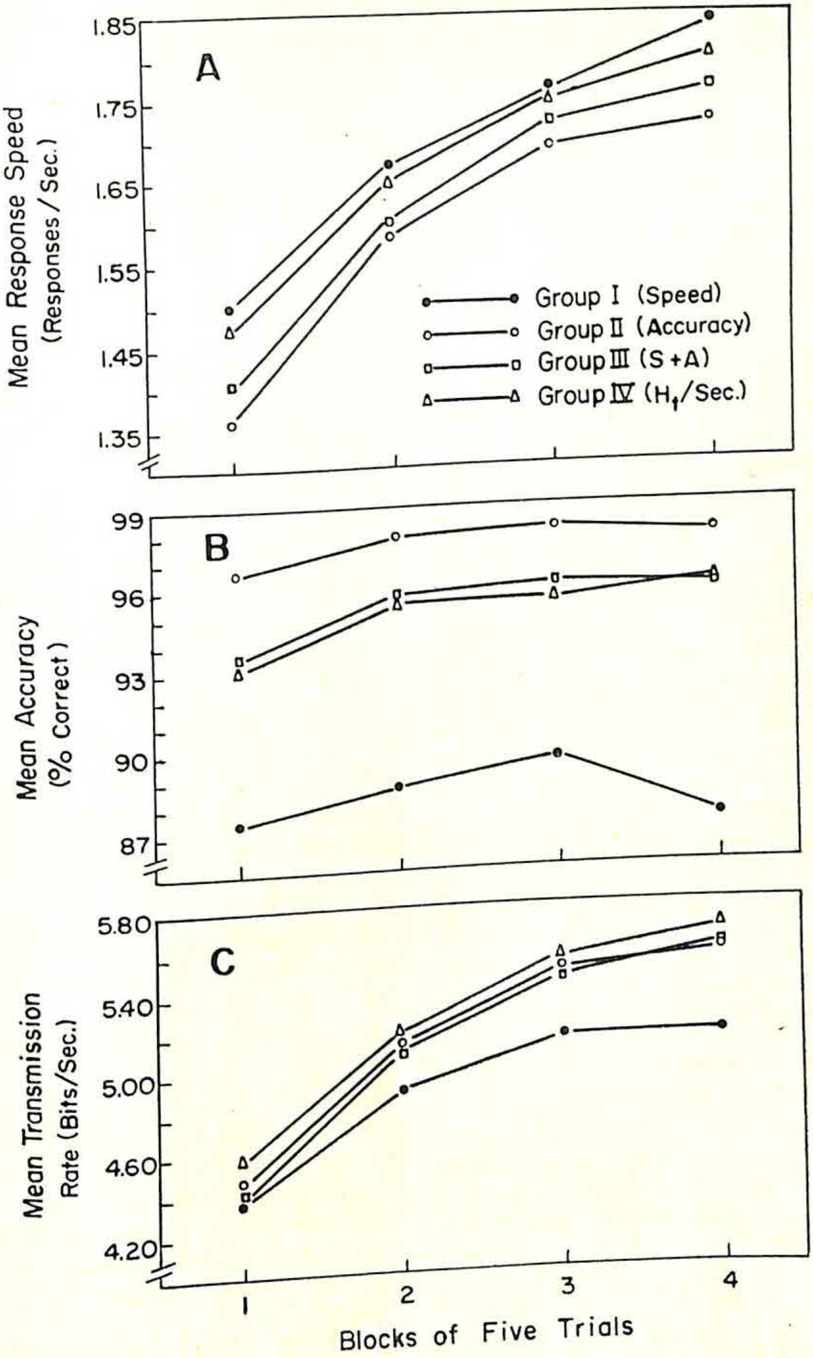


FIG. 1. Mean acquisition (Blocks 1–3) and performance (Block 4) scores for the three response indices: (a) speed, (b) accuracy, and (c) rate of information transmitted (H_t/sec).

TABLE 1

Comparison of Means and *SD*s for the Last Block of Trials on All Three Measures

Measure	Group I		Group II		Group III		Group IV	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Accuracy (% correct)	87.30	5.44	97.80	.41	95.90	.62	95.90	.75
Speed (response/sec)	1.84	.46	1.72	.26	1.76	.30	1.80	.40
Transmission Rate (H_t/sec)	5.16	.76	5.56	.40	5.58	.45	5.67	.60

It was found that all group differences were significant ($p < .05$) with the exception of that between Group III (speed and accuracy instructions) and Group IV (H_t/sec instructions).

The most critical results are those obtained using the H_t/sec metric, for it is this score which typifies most information-processing research. These data, which appear in Fig. 1c, indicate that performance was similar under all conditions save those for Group I (speed instructions). A Bartlett's test carried out on the performance data revealed a lack of heterogeneity among the variances of the four groups ($\chi^2 = 4.66$) which permitted the use of the F statistic to test group differences. The analysis yielded $F(3, 36) = 1.48$ which failed to achieve statistical significance. Since instructions do not constitute a continuous variable, and since only one group yielded scores which were at all deviant from the others, the t statistic was used to test the difference between Group I and all other groups combined. Viewed in this manner, the difference achieved significance at the level of $p < .05$.

Discussion

The similarity of Group III and Group IV performance on all three measures (speed, accuracy, and H_t/sec) suggests that contradictory instructions present no problem to *S* in a simple information-processing task. He appears, in fact, to weight the conflicting goals of speed and accuracy exactly as he would if he were attempting to maximize rate of transmission (H_t/sec). Furthermore, the fact that individual differences are no greater for Group III than for Group IV indicates that *S*s tend to agree in their weighting of speed and accuracy. These findings serve to strengthen the acceptability of investigations carried out under contradictory speed-accuracy instructions, particularly those studies in which performance is measured in terms of speed, accuracy, or H_t/sec.

With respect to the question of how multigoal performance compares with that for speed-emphasis or accuracy-emphasis instructions alone, the answer is clear and rather dramatic: *S* is primarily concerned with accuracy; he adjusts his response rate so that the percentage of errors does not exceed some critical level. In the present study the critical level appears to have been about 4% for both the speed-accuracy and H_t/sec conditions, a figure which is in agreement with some reported by other investigators (e.g., Garvey, 1957; Klemmer, 1958). Comparison of this accuracy level with those for Group I (speed) and Group II (accuracy) indicates closer agreement with the latter (2% errors) than with the former (13% errors). The high degree of similarity among speed-accuracy, H_t/sec, and accuracy group performance is corroborated by the response speed (Fig. 1a) and the information transmission (Fig. 1c) data.

Strong evidence for the notion of a critical accuracy level or accuracy criterion is provided by the acquisition functions shown in Fig. 1a and 1b. It is apparent here that accuracy remains at a fairly constant level throughout learning; nearly all improvement comes about as a result of increased response speed. This suggests that *S* is unwilling to lower his accuracy criterion even though it necessitates relatively slow responding. It is interesting to speculate upon the generality of this finding and to ask how rigidly *S* will maintain his accuracy criterion in the face of more difficult task requirements. Obviously, the answer to this question cannot be found in the present data since the task used was of maximum S-R compatibility. Comparison with other investigations (e.g., Alluisi & Muller, 1958; Klemmer, 1958), however, suggests that for less compatible tasks *S* reduces his accuracy criterion to levels as low as 89%. The precise nature of the relationship between *S*'s accuracy criterion and task difficulty or S-R compatibility is worthy of independent investigation.

Perhaps the most striking feature of the present data is the differential influence that speed-emphasis and accuracy-emphasis instructions have upon performance. It appears from Fig. 1b that accuracy emphasis does, in fact, lead *S* to raise his accuracy criterion slightly (from 96% in Groups III and IV to 98% in Group II), but with a negligible and statistically insignificant reduction in response speed (Fig. 1a). On the other hand, speed emphasis results in a negligible and insignificant increase in response speed and a very large reduction in the accuracy criterion (to 87% in Group I). It would appear, therefore, that in the absence of instructions to the contrary *S* behaves with the maximum efficiency permitted by his perceptual-motor equipment. He chooses a high accuracy criterion and responds as fast as this criterion will allow. He is, in all probability, capable of going faster, but the loss in efficiency incurred by such a choice is great (see Fig. 1c). Since his accuracy criterion is already high (at least for the present simple task), he cannot perform with much greater accuracy, and thus instructions to do so have a relatively small effect. It is interesting to note that *S* does not choose a 100% accuracy criterion even though strongly advised to do so by the accuracy-emphasis instructions. Apparently the loss in speed which such an approach would have incurred at the stage of training sampled was greater than *S* was willing to tolerate. Thus, it appears that there is a critical speed, as well as a critical accuracy level, below which *S* is unwilling to perform. This suggestion is implicit in the notion, expressed earlier, that *S*'s accuracy criterion is at least partially determined by S-R compatibility; if accuracy is sacrificed in a low-compatibility task, it is probably to avoid falling below some critical speed level.

Several conclusions may be drawn from these findings. First, the fact that instructions contain contradictory goals does not necessarily lead to serious variability among *S*s or to systematic biases in their mode of responding. This is not to question the cogency of the point raised by Edwards (1961); there may indeed be task situations in which contradictory instructions influence performance, but the typical information-processing schema does not seem to be among them. Second, it appears that it is feasible to use as a single common goal the metric of H_t/sec. Neither the data from the present study nor the informal questioning of *S*s after it suggest that there was any difficulty in understanding the concepts involved. Since scoring is generally based upon the information measure, it seems unnecessary to convert all values to the questionable money scale. Finally, it appears that *S* adopts certain speed and accuracy criteria in performing an information-processing task, which criteria are little influenced by instructional sets other than those emphasizing speed exclusively. Evidence from this and other studies suggests that the accuracy criterion chosen by *S* is high (90% or better) and that speed is adjusted to or near the highest level permitted by this criterion. However, it appears that there is a limit below which *S* will not adjust his speed; faced with a task which would force speed below this criterion, *S* chooses instead to lower his accuracy criterion. Much additional research is needed to establish the precise quantitative nature of these criteria, their relationships to each other and to specific task variables.

REFERENCES

ALLUISI, E. A., & MARTIN, H. B. An information analysis of verbal and motor responses to symbolic and conventional arabic numerals. *J. appl. Psychol.*, 1958, **22,** 79–84.

ALLUISI, E. A., & MULLER, P. F., JR. Verbal and motor responses to seven symbolic visual codes: A study in S-R compatibility. *J. exp. Psychol.*, 1958, **55,** 247–254.

ASHLEY, W., HARPER, R., & RUNYON, D. The perceived size of coins in normal and hypnotically induced economic states. *Amer. J. Psychol.*, 1951, **64,** 564–574.

ATTNEAVE, F. *Applications of information theory to psychology.* New York: Holt, 1959.

EDWARDS, W. Costs and payoffs are instructions. *Psychol. Rev.*, 1961, **68,** 275–284.

GARVEY, W. D. Operator performance as a function of the statistical encoding of stimuli. *J. exp. Psychol.*, 1957, **54,** 109–114.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4,** 11–26.

HOWELL, W. C. On the heterogeneity of stimulus and response elements in the processing of information. *J. exp. Psychol.*, 1962, **63,** 235–243.

KLEMMER, E. T. Time sharing between frequency-coded auditory and visual channels. *J. exp. Psychol.*, 1958, **55,** 229–235.

MOWBRAY, G. H., & RHODES, M. V. On the reduction of choice reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11,** 16–23.

(Received January 20, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 47–51

VERBAL PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF GROUPING SIMILAR STIMULI OR RESPONSES[1]

IRIS C. ROTBERG AND MYRON WOOLMAN[2]
Human Resources Research Office, George Washington University

Verbal paired-associate learning was measured when similar or dissimilar stimuli were grouped, and when similar or dissimilar responses were grouped. The following measures were employed: number of correct responses; type of errors made, i.e., errors indicating confusion between similar items and those indicating confusion between dissimilar items. The results indicated that learning was better when groups of stimuli were composed of similar items rather than dissimilar ones. The findings were interpreted in terms of discrimination and coding of the similar items.

The ability of *S*s to learn to associate pairs of items was measured (*a*) when similar or dissimilar items were grouped, and (*b*) when the similar or dissimilar groups of items were used either as stimuli or responses.

Gibson's (1942) research demonstrated that confusions between stimuli are a function of the degree of similarity between the stimuli. Subsequent studies have indicated that the manner in which similar stimuli are grouped affects the ease with which they are learned. In this connection, Gagné (1950) compared similar and dissimilar stimulus groups. Using nonsense forms as stimulus material, Gagné found that *S*s given learning groups composed of similar stimuli did better during testing than did those given dissimilar groups. He interpreted the finding in terms of the hypothesis that similar learning groups provide an earlier opportunity than do dissimilar groups for learning the cue relevant to the response. For the similar groups, earlier stimulus generalization is followed by relatively rapid discrimination and, therefore, superior test performance occurs.

Because of the nature of the response words that he used, Gagné could not demonstrate, for the majority of items, whether the errors resulted from confusion between similar items or dissimilar items. Therefore, his study used, as a measure of generalization, the number of overt errors, regardless of the type of confusion errors occurring.

It should be noted that Rothkopf (1958), using Morse Code stimuli, found the dissimilar groups superior. As a possible explanation, he noted that the stimuli in his study, unlike Gagné's, were not amenable to verbalization.

In the present study, the following measures were employed: (*a*) number of correct responses; (*b*) type of errors made, i.e., errors indicating confusion between similar items and those indicating confusion between dissimilar items. Each of these measures was compared for (*a*) *S*s given groups of stimuli composed of similar or dissimilar items and (*b*) *S*s given groups of responses composed of similar or dissimilar items.

---

[1] The research reported here was conducted by the authors while they were employed by the Human Resources Research Office, George Washington University, operating under contract to the Department of the Army. Opinions and conclusions are those of the writers and do not necessarily represent views of the university or the Department of the Army. The authors wish to thank their colleagues at HumRRO for their valuable comments and suggestions.

[2] Now at the Institute of Educational Research, Washington, D. C.

METHOD

Three experiments were conducted. In Exp. I, each stimulus and each response presented were displayed for 4 sec. To allow more time for responding, each stimulus and response were displayed for 8 sec. in Exp. II and III. Experiments I and II used the same operational definition of similarity. Experiment III used a second definition.

Subjects.—The *S*s were 154 male and female high school student volunteers from the District of Columbia. Fifty-six *S*s participated in Exp. I, 55 in Exp. II, and 43 in Exp. III.

Word list.—The nine-word paired-associate list of nonsense and English words and the organization of the list into similar and dissimilar groups are illustrated in Table 1.

The list of nine nonsense terms consisted of three similarity categories. The three words within each category had three letters in common (nondistinctive letters) and one letter differing (distinctive letter). No letter in any category was repeated in any other category.

Two types of similarity were employed. For Exp. I and II, nonsense terms were used in which the first three letters of each term within categories were nondistinctive letters. As may be noted from Table 1, these letters were identical in content and sequence. The fourth letter of each term was the distinctive letter. For Exp. III, the nonsense terms used the same three nondistinctive letters within each category. However, the letters differed systematically in sequence and position. The distinctive letter also varied in position.

TABLE 1

SIMILAR AND DISSIMILAR LEARNING GROUPINGS

Exp. I and II		Exp. III	
Similar	Dissimilar	Similar	Dissimilar
VKIW-YARD	VKIW-YARD	VKIW-YARD	VKIW-YARD
VKIH-STAR	LDAQ-MEAT	HVIK-STAR	DLAQ-MEAT
VKIF-ROAD	BJER-DOOR	KFIV-ROAD	BJER-DOOR
LDAQ-MEAT	VKIH-STAR	DLAQ-MEAT	HVIK-STAR
LDAT-TREE	LDAT-TREE	LTAD-TREE	LTAD-TREE
LDAX-COAT	BJEG-BOAT	XDAL-COAT	GBEJ-BOAT
BJER-DOOR	VKIF-ROAD	BJER-DOOR	KFIV-ROAD
BJEG-BOAT	LDAX-COAT	GBEJ-BOAT	XDAL-COAT
BJEN-HOUR	BJEN-HOUR	JNEB-HOUR	JNEB-HOUR

Words were combined for learning into three-word groups composed of either similar or dissimilar nonsense terms. The nonsense terms were randomly paired with common four-letter English words.

Experimental design.—Similarity was manipulated only in the nonsense term. Therefore, stimulus similarity was studied under a paired-associate procedure with English responses. Response similarity was studied under a procedure with nonsense responses.

The *S*s in each experiment were assigned to one of the following learning groups:

1. Similar Stimulus group—grouped presentation of similar stimuli (example, VKIW, VKIH, VKIF).
2. Dissimilar Stimulus group—grouped presentation of dissimilar stimuli (example, VKIW, LDAQ, BJER).
3. Similar Response group—grouped presentation of stimuli for which the correct responses are similar (example, YARD, STAR, ROAD).
4. Dissimilar Response group—grouped presentation of stimuli for which the correct responses are dissimilar (example, YARD, MEAT, DOOR).

Procedure.—For each experiment, *S*s were randomly assigned to treatment, and treatments were randomly assigned to day and hour. The same *E* and room were used for each experimental group. The word lists were presented by means of filmstrips projected on a screen facing the group.

First, a stimulus (nonsense or English) was flashed on the screen. While the stimulus was on the screen, *S*s were to write the response (English or nonsense) which they considered correct. The correct response was shown immediately after the presentation of the stimulus. The *S*s were directed to write their responses on every trial except the first trial of each three-word group.

During learning, each three-word group was presented for five trials. Words were rerandomized within groups each time they were repeated. The five trials for each three-word group were completed before the next group was presented.

Following learning, *S*s were given a test consisting of the nine words presented in random order for five trials. During the test period, correct responses were not shown.

TABLE 2

Mean Number of Correct Responses for Each Experimental Treatment during Testing

Exp.	Stimulus Groups		Response Groups	
	Similar	Dissimilar	Similar	Dissimilar
I	19.17	10.93	4.93	6.53
II	28.57	20.79	23.43	18.15
III	37.40	27.44	13.08	13.91

Results

The mean number of correct responses during testing is presented in Table 2 for each experimental treatment. The differences were evaluated by means of two-tailed *t* tests.

The Similar Stimulus groups made more correct responses than did the Dissimilar Stimulus groups. The difference was significant in Exp. III ($p < .02$), and approached the .05 significance level in Exp. I ($p < .057$) and Exp. II ($p < .051$). Differences between the Similar and Dissimilar Response groups were not significant.[3]

Errors were classified into inside-category (similar) and outside-category (dissimilar) substitutions. In the stimulus groups, an inside substitution is defined as the use of a response word appropriate to a term from the same similarity category. An outside substitution is the use of a response word appropriate to another similarity category. For example, if the stimulus is VKIW, an inside substitution is STAR or ROAD, and an outside substitution is MEAT, TREE, COAT, DOOR, BOAT, OR HOUR. (See Table 1.)

For the response groups, an inside substitution is defined as the substitution of a letter from the same similarity category. An outside substitution is the substitution of a letter from another category. The measure used is number of erroneous occurrences of the distinctive letters, regardless of position. For example, if the stimulus is YARD, the response KFBT includes one inside substitution (F) and one outside substitution (T).

The mean number of errors of each type for each experimental group is presented in Table 3.

---

[3] It can be noted from Table 2 that the groups whose responses were English words made more correct responses than did the groups whose responses were nonsense terms. These differences were significant for the similar ($p < .01$) and dissimilar ($p < .02$) groups in Exp. I and for the similar ($p < .001$) and dissimilar ($p < .02$) groups in Exp. III.

TABLE 3

Mean Number of Inside and Outside Substitution Errors for Each Experimental Treatment during Testing

Exp.	Substitution Errors	Stimuli				Responses			
		Similar		Dissimilar		Similar		Dissimilar	
		Weighted Mean	Raw Mean	Weighted Mean	Raw Mean	Weighted Mean	Raw Mean	Weighted Mean	Raw Mean
I	Inside	20.76	6.92	39.99	13.33	13.92	4.64	23.19	7.73
	Outside	14.42	14.42	16.47	16.47	10.00	10.00	12.73	12.73
II	Inside	20.13	6.71	29.79	9.93	25.50	8.50	22.62	7.54
	Outside	9.00	9.00	13.29	13.29	7.50	7.50	13.38	13.38
III	Inside	9.30	3.10	19.32	6.44	37.38	12.46	21.54	7.18
	Outside	3.70	3.70	9.56	9.56	16.54	16.54	14.64	14.64

The inside substitutions were weighted by 3 because their chance probability of occurrence was one-third the probability of outside substitutions. The raw number of substitutions for each treatment is also presented in the table.

Comparisons indicate the following significant differences:

Within stimulus groups.—More inside (weighted by 3) than outside substitution errors were made by the Similar ($p < .05$) and Dissimilar ($p < .01$) Stimulus groups in Exp. I and II, and by the Dissimilar Stimulus group in Exp. III ($p < .05$).

Across stimulus groups.—The Similar Stimulus group made fewer inside substitutions than did the Dissimilar Stimulus group in Exp. I ($p < .01$), and fewer outside substitutions than did the Dissimilar Stimulus group in Exp. III ($p < .05$).

Within response groups.—More inside (weighted by 3) than outside substitutions were made by the Dissimilar Response group in Exp. I ($p < .01$), by the Similar ($p < .02$) and Dissimilar ($p < .01$) Response groups in Exp. II, and by the Similar Response group in Exp. III ($p < .001$).

Across response groups.—The Similar Response group made more inside substitutions than did the Dissimilar Response group in Exp. III ($p < .01$).

DISCUSSION

The results of this study show that learning is better when groups of stimuli are made up of similar items rather than dissimilar ones. As previously described, Gagné (1950) suggested that groups of stimuli consisting of similar items provide a discrimination advantage. This explanation would be sufficient in the present study if the Similar Stimulus groups were superior only because they made fewer inside errors. However, this explanation does not account for the fact that the Similar Stimulus groups made fewer outside errors as well, in Exp. II and III. Only in Exp. I can the overall superiority of the Similar Stimulus group be attributed primarily to fewer inside errors.

We suggest that the Similar Stimulus groups made fewer outside errors for the following reasons: Similar stimulus grouping provides stimulus categories that can be learned readily. The responses paired with the stimuli within categories will also, to some degree, be associated with each other. During testing, *S* can assign each stimulus to one of the categories. To the extent that he has also linked responses to each other and to the similarity category, *S* can reduce the number of alternatives from which he must choose. Thus, for example, although there are nine possible responses to the stimulus VKIW, *S*s from the Similar Stimulus group are more likely to limit their responses to YARD, STAR, or ROAD—all of which belong to the appropriate similarity category. When dissimilar stimuli are grouped, stimuli belonging to similarity classifications are separated and, therefore, no easily identifiable categories are available.

Rothkopf (1958) explained the superior performance of the Dissimilar Stimulus group in his study by noting that *S*s were unable to verbally describe the stimuli (Morse Code). The apparent inconsistency between these results and the results of Gagné (1950) and those cited in the present paper may be in part resolved by noting that the inability to verbalize might reduce the coding advantage provided by the Similar Stimulus group.

In summary, stimulus similarity results in learner generalization, manifested both in discrimination difficulty and coding opportunity. Thus, similar stimulus grouping is superior to dissimilar grouping because it decreases discrimination difficulty, while increasing the opportunity for coding. Within the present research structure, an inability to discriminate has been measured in terms of *S*s' confusions between similar items (inside substitution errors). An inability to code, that is, an inability to assign responses to the proper similarity category of items, has been measured in terms of *S*s' confusions between dissimilar items (outside substitution errors).

Generalization between similar responses is evident from the error analyses. However, the present results do not indicate that grouping similar responses (as distinguished from similar stimuli, described above) is effective for error reduction.

Grouping similar responses apparently does not provide the possibility either for accelerated discrimination or for coding. Accelerated discrimination might not be evident because the role of the distinctive cue is not so sharply delineated. Coding probably does not occur because *S*s have not classified stimuli and, therefore, cannot assign them to similarity categories during testing.

We have hypothesized that discrimination and coding can operate most effectively when the grouped similars are in the stimulus position. The similar items can be functionally placed in this position by the appropriate use of stimulus and response correction procedures.

In the stimulus correction procedure, *S*s are shown the stimulus, make their responses, and then are shown the correct stimuli for the responses they made. This procedure functionally places responses in the stimulus position. Therefore, a stimulus correction method should be more effective than a response correction method for reducing errors under similar response grouping.

In the response correction procedure, *S*s are shown the stimulus, make their responses, and then are shown the correct response for the stimulus presented. This is the usual procedure and the one employed in the present experiments. A response correction method should be more effective under similar stimulus grouping (Shepard, 1958).

REFERENCES

GAGNÉ, R. M. The effect of sequence of presentation of similar items on the learning of paired associates. *J. exp. Psychol.*, 1950, **40**, 61–73.

GIBSON, E. G. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30**, 185–200.

ROTHKOPF, E. Z. Stimulus similarity and sequence of stimulus presentation in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 114–122.

SHEPARD, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509–523.

(Received January 20, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 52–58

CLASSICAL EYELID CONDITIONING AS A FUNCTION OF SUSTAINED AND SHIFTED INTERSTIMULUS INTERVALS[1]

HARVEY C. EBEL AND WILLIAM F. PROKASY

Pennsylvania State University

39 Ss were given 1200 classical eyelid conditioning trials over 12 sessions of 100 trials each. The manipulated independent variable was interstimulus interval, the specific values employed being 200, 500, and 800 msec. Depending upon the treatment combination, some Ss received only 1 interstimulus interval throughout training, while others were shifted back and forth between 2 interstimulus interval values. The major results were (a) response frequency continued to increase beyond 400 training trials, though differences as a function of interstimulus interval were not statistically significant; (b) both mean and SD of latency varied directly with interstimulus interval and changes in interstimulus interval; and (c) both mean and SD of latency decreased across training sessions. The data were interpreted in terms of an assumed reinforcement related to the temporal disparity between UCR and UCS onsets.

The results of a study by Boneau (1958) make it clear that latency, as well as frequency, of CR is influenced by manipulations of interstimulus interval (ISI). Boneau reported that latencies increase with ISI and that, as training progresses, they increase and become more precise with respect to UCS onset. These observations strongly suggest that a major component of what *S* learns in classical eyelid conditioning is when to make the response; i.e., CR latency.

It was the primary purpose of this study, therefore, to examine characteristics of CR latency in more detail through the manipulation of ISI. Emphasis was placed upon individual latency distributions and how they were modified by initial training to ISI "A," a shift to ISI "B", and a return to ISI "A." In order to examine complete individual distributions, however, it was necessary to use an intermittent reinforcement training procedure, so that post-UCS CRs could be included in the distributions.

METHOD

Apparatus.—The apparatus for picking up, amplifying, and recording eyelid responses is described elsewhere (Prokasy & Whaley, 1961).

The CS was an increase in brightness from 5.2 to 295.6 mL. of a 3-in. milk-glass disk located approximately 32 in. in front of *S*, about 11 in. below eye level. The UCS was a 50-msec. puff of nitrogen delivered to *S*'s right cornea through a 4.5-mm. diameter jet adjusted to between .25 and .75 in. from *S*'s eye. The puff intensity was sufficient to support a 270-mm. column of mercury, measured at point of delivery to the eye.

Trials were programed at a fixed intertrial interval (ITI) of 10 sec. through a Western Union tape transmitter. The CS, UCS, and ISI durations were controlled by Hunter-Brown interval timers. Equipment cues were masked by an 80-db. spl white noise sound screen presented through a set of earphones.

---

[1] This report is an adaptation from the master's thesis of Harvey C. Ebel, under the direction of William F. Prokasy, submitted to the Graduate School of the Pennsylvania State University.

The research was supported in part from a grant from the National Science Foundation (NSF-G-7463) and in part from a grant from the Council on Research at the Pennsylvania State University. The authors would like to acknowledge the aid of Dorsey Johnson in helping to run *S*s and compile data.

Subjects.—The *S*s were 46 male undergraduate students at the Pennsylvania State University, enrolled in a class in introductory psychology, who participated in the study to fulfill a course requirement. Of the 46 *S*s, 2 were eliminated because they inadvertently were placed under the wrong ISI condition after training procedures had been initiated, and another 5 were eliminated because they failed to give UCRs to the UCS.

Experimental design.—The 39 *S*s who completed training were placed randomly into one of nine treatment groups. Three ISI conditions were employed in the study: 200 msec., 500 msec., and 800 msec. Of the nine treatment combinations, six involved shifts from one ISI to another, while three involved only a single ISI value throughout training.

Five *S*s each were assigned to the six shift conditions, and 3 *S*s each to the control conditions. Each *S* underwent 12 sessions of 100 trials each. For *S*s in the six shift groups, the first four sessions (Block 1) were devoted to experience with the initial ISI, the second four (Block 2) to a different ISI, and the final four (Block 3) to the initial ISI. The first session in each block was an 80% reinforcement schedule with 20 test trials (i.e., nonreinforced trials) randomly inserted in the total of 100 trials. The following three sessions in each block were 65% reinforcement schedules, containing 35 test trials in the total of 100 trials. Thus, each block contained a total of 400 trials, and each *S* was to receive a total of 1200 trials during the course of the 12 sessions. The three consistent (i.e., control) groups received either a 200-, 500-, or 800-msec. ISI throughout the 1200 trials. To control for number of reinforced trials, *S*s in these conditions received the 80 and 65% schedules in the same order as that given to the six shift groups.

Procedure.—Neutral instructions were read to *S*s. They indicated the function of the various pieces of equipment within the testing chamber, and emphasized that *S*s were to watch the milk-glass disk, relax, and not control voluntarily their natural reactions to the stimuli. After the instructions were read to *S*, he was asked to repeat them "in general" and then the recording headgear was fitted. The *E* left *S*'s chamber and made any necessary apparatus adjustments, after which he initiated training trials.

The total time elapsing from beginning of instructions to completion of the 100 trials in the first session was approximately 25 min. Individual sessions were separated, generally, by no more than 4 days. A CR was defined as the first response on a test trial occurring in the interval from 166 to 1066 msec. after CS onset. Through extending the recording period beyond the 200-, 500-, and 800-msec. CS durations by, respectively, 866, 566, and 266 msec, this definition of a CR range equalizes the opportunity for involuntary blinks in the three ISI conditions. The CS duration equaled ISI in all conditions.

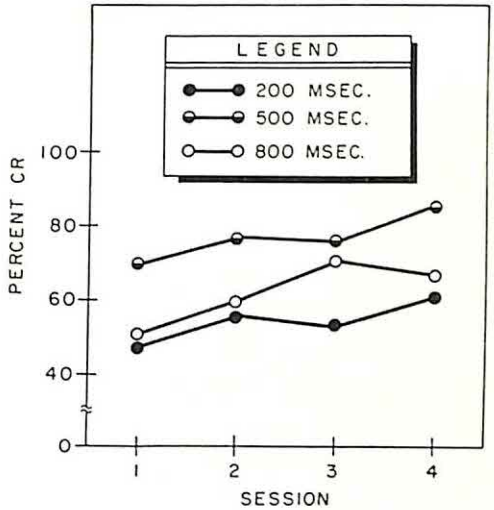


FIG. 1. Percent CR over the first four sessions as a function of ISI. (*N* is 13 in each group.)

RESULTS

Frequency.—For the first four sessions there were 13 *S*s each receiving an ISI of 200 (Groups 222, 252, 282),[2] 500 (Groups 555, 525, 585), and 800 (Groups 888, 828, 858) msec. Figure 1 provides mean percent CR on test trials for those four sessions and Table 1 summarizes an analysis of variance of the data. Although *S*s receiving the 500-msec. ISI performed consistently highest, the differences among groups were not significant.

In addition to the reliable individual

---

[2] Group designations correspond to block-by-block treatment. For instance, Group 252 received a 200-msec. ISI in Block 1, a 500-msec. ISI in Block 2, and a 200-msec. ISI in Block 3.

TABLE 1

Analysis of Variance of Percent CR as a Function of ISI and the First Four Training Sessions

Source	*df*	*MS*	*F*
Between *S*s	38		
ISI	2	689.9	2.1
Error (b)	36	328.5	19.1***
Within *S*s	117		
Sessions (S)	3	129.0	7.5***
ISI × S	6	20.6	1.2
Error (w)	108	17.2	

*** $p < .001$.

differences (significant b-error), percent CR continued to increase through the fourth session, with no evidence that asymptotic performance had been reached. That performance, in general, increased beyond the fourth session is suggested by the medians for each group provided in Table 2. Small group *N*s at this stage of practice precluded satisfactory statistical analyses of these data, though increases in median percent CR from Session 4 to Sessions 10–12 were observed in eight groups with one tie. Discounting the tie, this result is significant at the .01 level by a two-tailed binomial test.

Since initial group differences were not large and CR performance increased throughout sessions, there was no way to determine the effect of shifts in ISI on percent response.

TABLE 2

Median Percent CR on Session 4 and on Sessions 10–12

Group	Median % CR	
	Session 4	Sessions 10–12
222	50	100
252	50	83
282	83	90
555	100	100
525	80	93
585	93	97
888	50	83
828	63	87
858	90	93

Latency.—For each *S* three latency distributions, each based on 90 test trials,[3] were obtained: one from Sessions 2–4, one from Sessions 6–8, and one from Sessions 10–12. By omitting latencies from the 20 test trials on each of Sessions 1, 5, and 9, heterogeneity of the distributions was reduced, since these sessions would be influenced by initial acquisition effects and the immediate effects of shifts from one ISI to another. The data of 1 *S* in each of the nine groups are presented in Fig. 2 and 3.[4]

Several trends are evident in these distributions. First, omitting alpha responses (those responses with latencies under 166 msec.), the distributions tend to be unimodal and symmetric, with an increasing density around the mode as training progresses (Fig. 2). Second, these figures exemplify the control on latency exerted by the ISI: the longer the ISI, the greater the latency and the less dense the latency distribution. Third, there tends to be a decrease in latency (Fig. 2) across training blocks.

For each block of four sessions both mean and *SD* of latency were obtained for individual *S*s in all groups. The obtained group averages are provided in Table 3.

During Block 1, before ISI shifts were effected, there were 13 *S*s each

---

[3] Unpublished data from our laboratory reveal systematic latency increases over the first 20 trials of each session; therefore, the first 5 test trials, occurring early in each session, were omitted from analyses.

[4] Data for each individual *S*, including session-by-session frequencies and block-by-block latency distributions as well as the results of more extensive statistical analyses, are available upon request from the authors. Group trends reported in this paper held for the majority of individual *S*s.

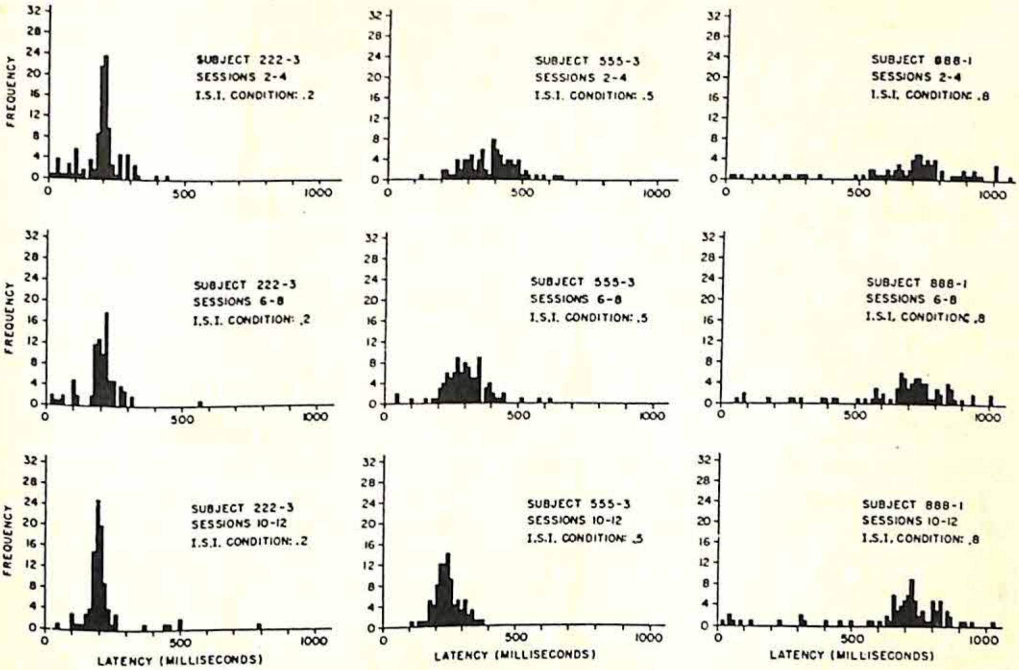


FIG. 2. Individual *S* latency distributions at three stages of training for 1 *S* each from Groups 222, 555, and 888.

receiving an ISI of either 200, 500, or 800 msec. The pooled mean latencies of, respectively, 311, 465, and 664 msec. were significantly different, $F(2, 36) = 33.9$; $p < .001$; error $MS = 12000$. Similarly, mean *SD*s of, respectively, 120, 128, and 193 msec. differed significantly, $F(2, 36) = 4.7$; $p < .05$; error $MS = 4818$.

TABLE 3

LATENCY MEANS AND *SD*s (MSEC.) OVER THREE BLOCKS OF FOUR SESSIONS

Group	Block 1		Block 2		Block 3	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
222	268	67	245	65	231	57
252	342	137	425	163	302	97
282	306	135	465	199	286	129
555	527	128	413	82	305	69
525	499	153	314	131	469	153
585	393	87	575	158	420	107
888	760	193	773	145	693	171
828	623	213	308	116	515	224
858	647	172	508	106	558	181

The significance of the latency decrease in Groups 222, 555, and 888 across blocks was tested with a Jonckheere (1954) test for ordered alternatives and related measures. The particular scores employed were individual means on each of the final 11 sessions. Despite the fact that there were only 3 *S*s in each group, the obtained p values for Groups 222, 555, and 888, respectively, were .006, $<.001$, and .063 ($z = 2.52$, 5.14, and 1.54).

For each of the total of 9 *S*s in Groups 222, 555, and 888 mean *SD*s on Blocks 1 and 3 were compared. A t test for related measures yielded a $p < .05$, $t(8) = 2.42$; $\sigma_m = 12.5$ for the mean decrease of 30.3 msec.

Table 3 illustrates how increases or decreases in ISI resulted in correlated increases or decreases in both mean and *SD* of latency. Of 30 *S*s in the six shift groups, the change in ISI from Block 1 to Block 2 resulted in a

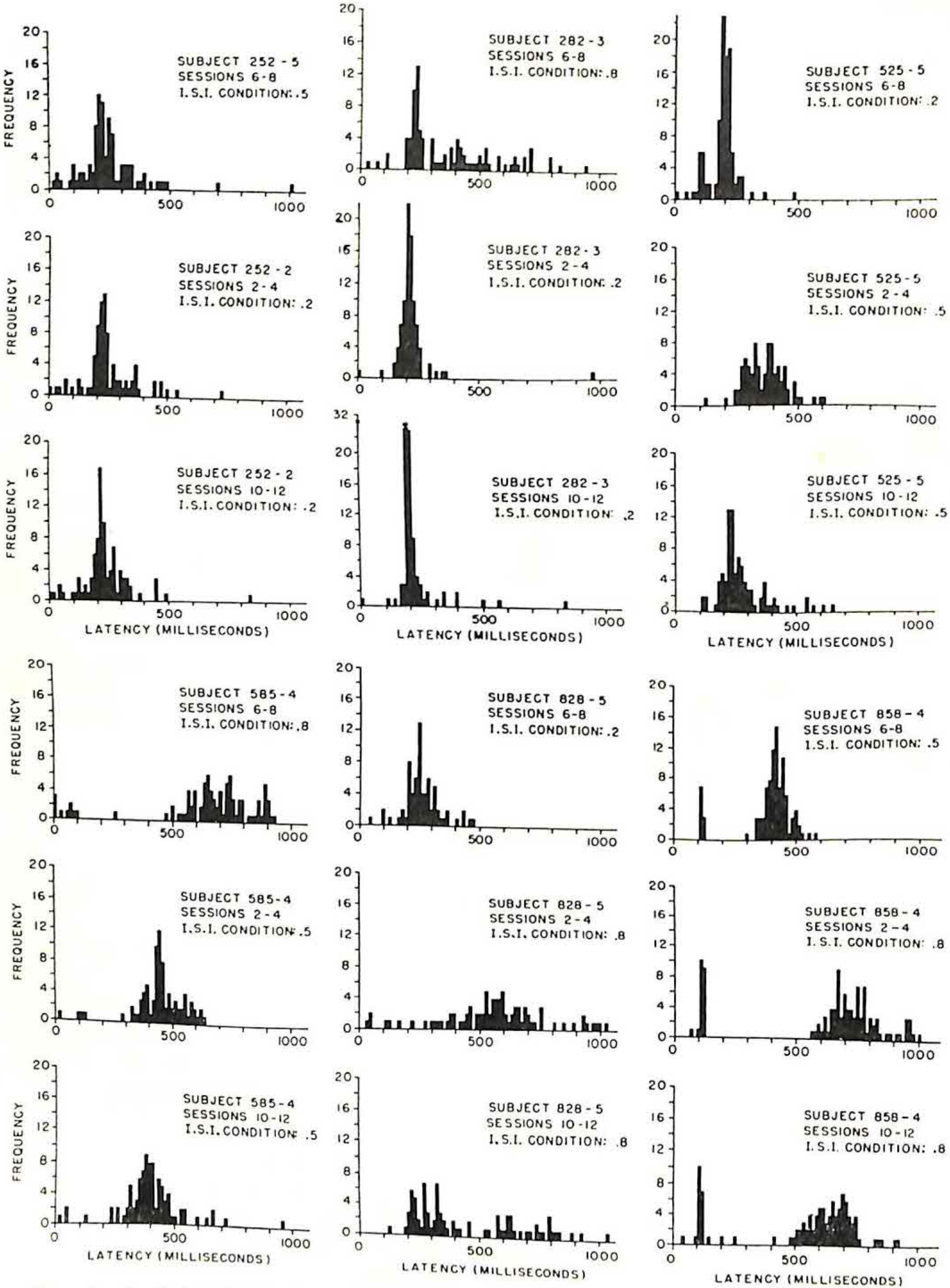


FIG. 3. Individual *S* latency distributions at three stages of training for 1 *S* each from Groups 252, 282, 525, 585, 828, and 858. (The first stage of training is presented as the middle distribution for each *S* in order to facilitate comparisons of first and third stages.)

corresponding increase or decrease in mean and *SD* with, respectively, 28 and 27 *S*s. Similarly, in the ISI change from Block 2 to Block 3, the means and *SD*s of 27 and 26 *S*s, respectively, corresponded with the direction of change. All four of these results are significant with a two-tailed binomial test ($p < .001$).

To some extent the overall latency decrease in Groups 222, 555, and 888 can be observed in the six shift groups. Specifically, a comparison of mean latencies between Blocks 1 and 3 was made and, of 30 *S*s, 22 yielded a shorter mean latency in Block 3, significant at beyond the .02 level with a two-tailed binomial test. It should be noted that three of the exceptions were in Group 585, two in Group 525, two in Group 282, and one in Group 252.

DISCUSSION

Since response frequency did not vary significantly as a function of ISI, our results provide no information on how within-*S* shifts in ISI affect this dependent variable. An unanticipated result, however, was that overall response frequency continued to increase beyond Training Trial 400, and that final median percent response was 83 or greater for all groups irrespective of the ISI conditions employed. Undoubtedly the use of 12 separate sessions made some contribution to the continued increase, since there is evidence (Prokasy, Grant, & Myers, 1958) of recovery between successive acquisition sessions. Nonetheless, with massed training trials (10-sec. intertrial intervals) and a low intermittent reinforcement percentage (65 on 9 of the 12 sessions), both of which retard CR development (see Kimble, 1961, pp. 103, 125), the high response levels obtained are somewhat surprising. Since past studies of the ISI variable in classical eyelid conditioning were conducted with fewer than 200 trials, our results raise the possibility that the ISI function does not reflect exclusively asymptotic performance.

Of major concern are the results that individual latency means and *SD*s were positively related to ISI and that within-*S* increases or decreases in ISI resulted in correlated changes in latency means and *SD*s. Both sets of observations are consistent with the group data of Boneau (1958). In addition, it was observed that individual latency distributions tended to be symmetric, with a portion of some distributions extending beyond the point in time at which the UCS was administered. The latter was particularly noticeable with an ISI of 200 msec., where all latency means exceeded 200 msec. There was some suggestion that asymptotic mean latency with a particular ISI was influenced only negligibly by prior experience with other ISIs, although adequate tests for this were precluded since there was an overall latency decrease throughout the 1200 training trials.

By applying Logan's (1956) micromolar analysis to eyelid conditioning, Boneau (1958) interpreted different CR latencies as different responses, with reinforcement presumed to have occurred when latency was sufficiently short to result in puff avoidance. While our data do not permit the deduction that blinks of different latencies are different responses, we agree with Boneau that reinforcement in classical eyelid conditioning is contingent, in part, upon the amount of time separating the CR and UCS events. The temporally appropriate location of the CR with respect to UCS onset and the increased density of individual latency distributions resulting with extensive training strongly suggest a kind of instrumental shaping in which the kinesthetic and tactual feedback from each CR-UCS outcome plays an important role. Such an interpretation, it should be noted, is consistent with the theoretical interpretations of conditioning provided by Birch and Bitterman (1949), Perkins (1955), and Jones (1961).

If the assumption is made that the eyelid response is shaped through a reinforcement contingency based in part on the CR-UCS interval, then an interpretation of the function relating CR frequency and ISI can be made: the ISI function should reflect the efficiency with which *S* is able to maximize the CR-UCS reinforcement contingency. As ISIs increase beyond 500 msec., the accompanying increase in the variance of *S*'s latency distribution should result in a reduction of reinforcement, and, consequently, a reduction in response frequency. Though the variance of the latency distribution is small with short (200 msec.) ISIs, the effectiveness of a CR will be reduced since *S*s are apparently unable to make an eyelid CR with a latency of much under 180 msec. Consistent with past data (Kimble, 1947), therefore, frequency of CR would be expected to reduce rapidly with ISIs shorter than 200 msec. The optimum range of ISIs would be that in which it is possible for *S* to make a CR before UCS onset and also in which it is possible for *S* to make the CR with a rather precise time discrimination.

REFERENCES

BIRCH, H. G., & BITTERMAN, M. E. Reinforcement and learning: The process of sensory integration. *Psychol. Rev.*, 1949, **56**, 292–308.

BONEAU, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

JONCKHEERE, A. R. A test of significance for the relation between m rankings and K ranked categories. *Brit. J. statist. Psychol.*, 1954, **7**, 93–100.

JONES, J. E. The CS-UCS interval in conditioning short- and long-latency responses. *J. exp. Psychol.*, 1961, **62**, 612–617.

KIMBLE, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–15.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* (2nd ed.) New York: Appleton-Century-Crofts, 1961.

LOGAN, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63**, 63–73.

PERKINS, C. C., JR. The stimulus conditions which follow learned responses. *Psychol. Rev.*, 1955, **62**, 341–348.

PROKASY, W. F., JR., GRANT, D. A., & MYERS, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

PROKASY, W. F., & WHALEY, F. L. The intertrial interval in classical conditioning. *J. exp. Psychol.*, 1961, **62**, 560–564.

(Received January 27, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 59–62

A TEST OF THE "UNITS HYPOTHESIS" EMPLOYING WAVE-LENGTH GENERALIZATION IN HUMAN SUBJECTS[1]

DAVID R. THOMAS AND RICHARD H. HISS[2]

Kent State University

An experiment was performed to test 2 hypotheses suggested by Mednick and Freedman (1960) concerning the role of the number of intervening test stimuli in determining generalization decrement. 3 different groups of 20 female college students viewed a monochromatic light of 530 mμ. (green) for 60 sec. and then were subjected to 9 different random series of wave lengths, with instructions to respond only to the original color. Group I was tested for generalization with the series 520–540 mμ. in 5-mμ. steps; Group II received 510–550 mμ. in 10-mμ. steps; whereas Group III received 490–570 mμ. in 20-mμ. steps. In agreement with Mednick and Freedman it was found that the more stimuli intervening between the CS and the test stimulus the greater the generalization decrement. Contrary to their position, however, the physical size of the stimulus unit was also found to be a significant factor.

Early experiments in the area of stimulus generalization (e.g., Gibson, 1939; Hovland, 1937; Humphreys, 1939) often employed stimulus values selected on the basis of a jnd scale. On the other hand most of the more recent studies have employed physical stimulus dimensions, measured directly in terms of size, brightness, wave length, or similar units. This trend continues despite the fact that many recent studies have shown that, at least as far as human *S*s are concerned, physical stimulus differences alone are not sufficient to account for generalization effects. An example is provided by the work of Mednick and Freedman (1960) who have shown the number of stimuli employed in testing for generalization to be an important determiner of generalized response strength. They postulate that the amount of generalization is proportional to the number of stimulus units that separate the test stimulus from the original one. On this basis, Mednick and Freedman suggest that (*a*) for a given physical distance from the CS the greater the number of intervening stimuli employed the greater the generalization decrement, and (*b*) a stimulus value one (two, three, etc.) units removed from the CS will produce the same decrement no matter what the physical size of the unit.

The purpose of the present study was to test these two hypotheses employing human *S*s in a wave-length generalization situation. The procedure employed was that originally developed by Kalish (1958) and was subsequently used in a number of studies in our laboratory (Thomas & Jones, 1962; Thomas & Lanier, 1962; Thomas & Mitchell, 1962).

METHOD

Subjects.—The *S*s were 60 undergraduate women selected from introductory psychology courses at Kent State University.

---

[1] This study was supported in part by Grant M-4203 from the United States Public Health Service, National Institute of Mental Health. The authors would like to acknowledge the contribution of Herbert R. Friedman of Duke University toward the conceptualization of this experiment.

[2] Now at Syracuse University.

Apparatus.—The study employed a Skinner-type key pecking apparatus, modified for use with human *S*s. The box was approximately 15 in. long, 14 in. wide, and 11 in. high, and was painted flat black. The front wall of the box was made of transparent Plexiglas so that *S*s could clearly view the pecking key. The *S* sat in a chair approximately 2½ ft. from the key, which was a circular plastic disc ⅞ in. in diameter. Illumination was provided by a Cambridge Thermionic Corporation monochromator, Model B Series 1066, equipped with an Olympus Model 201250 6-v. 5-amp. light source. The patch of color on the key was approximately 4 mL. in luminance. The only other light in the room was a 7.5-w. "night light" on *E*'s side of a black cloth screen separating *S* from *E*. A telegraph key was used as the manipulandum. This key was placed on the table next to the experimental box, within easy reach of *S*'s right hand. It was wired so that its release would activate a signal light on *E*'s side of the screen, thus signifying a response. The box was equipped with an electrically operated shutter which interrupted the monochromator beam when *E* threw a switch.

Procedure.—The 60 *S*s were unsystematically divided into three groups of 20 *S*s each. The treatment of the three groups was similar with the exception that different sets of values were used in the test for generalization. All *S*s received the following instructions:

> This is an experiment in color perception. At the beginning of the experiment a specific color will be presented through the small hole in front of you. Try to keep this color in mind because you will be asked to identify it later. After 1 min. this color will be turned off and you will place your finger and press down on the telegraph key in front of you. I will give the signal "ready" and a few seconds later a color will be again presented. You must decide whether this is the original color shown you at the start of the experiment. If it is, *lift* your hand as rapidly as you can from the key. If it is not, keep pressing on the key.
>
> I will say the word "ready" whenever I am about to present a color and you should be pressing the key at that time. We are going to try some practice trials. Now we are going to run through a series exactly as we would if this were the real experiment.

At this point *E* exposed 600 mμ. for 60 sec. and then presented 610, 590, 600, and 580 mμ.

> Now we are going to begin the experiment. Remember, try to keep the original color in mind and respond as quickly as you can, *lifting* your finger only when the original color appears. Any questions?

The *E*'s reading of the instructions to *S* was interrupted at the designated time for the practice series. If *S* appeared to have understood the instructions and had no further questions, the standard stimulus (530 mμ., a middle green) was exposed for 60 sec. Then nine randomized series of four test stimuli plus the standard were presented. Each test stimulus was presented for a period of 3 sec. and was followed by a blackout of 5 to 10 sec., allowing sufficient time for *E* to record the presence or absence of a response and change the monochromator setting.

For each group five test stimuli were used, the original stimulus and two wave-length values on each side of the original. For Group I the test stimuli were 520, 525, 530 (the standard stimulus), 535, and 540 mμ. Thus, the size of the stimulus unit for Group I was 5 mμ. For Group II the size of the stimulus unit was 10 mμ. The values used were 510, 520, 530 (the standard stimulus), 540, and 550 mμ. For Group III the stimuli were spaced 20 mμ. apart. The test values 490, 510, 530 (the standard stimulus), 550, and 570 mμ. were employed.

RESULTS AND DISCUSSION

The generalization gradients of the three experimental groups are shown in Fig. 1. These curves were generated by plotting the mean number of responses made during the generalization test against the wave length of the test stimuli. It may be seen that all groups show decremental gradients with greatest response strength to the original stimulus value. The relative importance of wave length and number of stimulus units may be determined by inspection of Fig. 1. If wave length were the sole determiner of generalized response strength the three curves would coincide for those stimulus values for which measures are available for comparison. The four stimulus values 510, 520, 540, and 550 mμ. are the ones of interest. In this case of all four values a comparison may be made between the mean number of re-

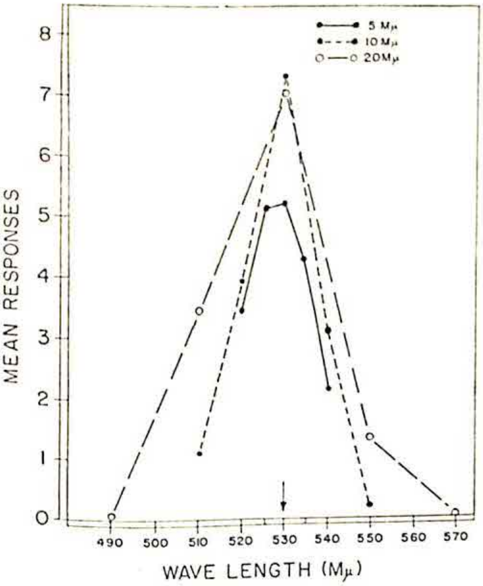


FIG. 1. Stimulus generalization as a function of wave length with size of stimulus unit as parameter.

sponses emitted by two groups, each experiencing a different number of stimulus units between that value and the standard stimulus value.

For the stimulus value 510 mμ. the mean number of responses given by the 10-mμ. group was 1.10 compared to 3.45 for the 20-mμ. group. This difference is significant, $t(38) = 2.72$, $p < .01$. For the value 520 mμ. the 5-mμ. group gave a mean of 3.45 responses compared to a mean of 3.95 responses for the 10-mμ. group. For the value 540 mμ. the 5-mμ. group gave a mean of 2.15 responses compared to 3.10 for the 10-mμ. group. These differences, although in the predicted direction, fail to achieve statistical significance. For the stimulus value 550 mμ. the 10-mμ. group gave 0.15 responses compared to 1.30 responses in the 20-mμ. group. This difference is significant, $t(38) = 2.88$, $p < .01$. Thus, all four comparisons are in the expected direction; the greater the number of stimulus units a given wave length is removed from the original stimulus, the less the generalized response strength.

The data presented in Fig. 1 indicates that the number of stimulus units involved must be considered in the prediction of generalized response strength. The suggestion by Mednick and Freedman (1960) that the number of stimulus units is the sole determiner of generalized response strength may best be evaluated with reference to Fig. 2. In this figure, mean number of responses is plotted as a function of the number of stimulus units from the standard, with size of stimulus unit as parameter. If the size of the stimulus unit made no difference, that is, if the number of stimulus units were the sole determiner of generalized response strength, then the three curves would coincide. Clearly this is not the case.

The data in Fig. 2 were subjected to an analysis of variance. The result was a significant F for unit size, $F(2,57) = 17.89$, $p < .01$; number of

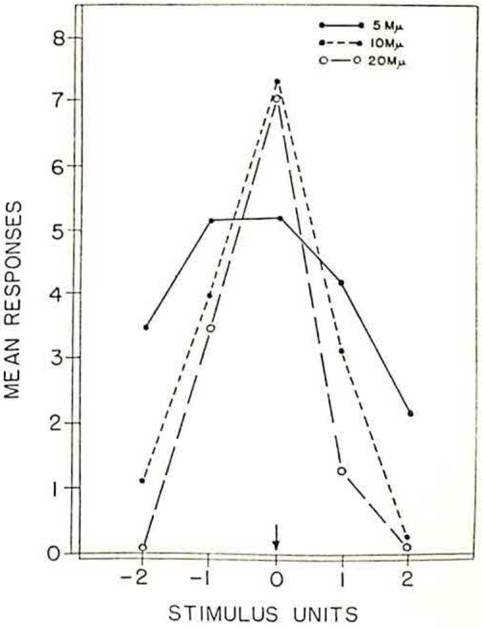


FIG. 2. Stimulus generalization as a function of number of stimulus units, with size of stimulus unit as parameter.

stimulus units, $F(4,228) = 47.40$, $p < .01$; and Number of Units × Size of Unit interaction, $F(8,228) = 4.09$, $p < .01$. Of particular interest is the significant interaction between number and size of stimulus units. This indicates that the number of responses given to a stimulus one unit removed (or two) from the standard varies with the physical size of the unit. It should be noted that the F for stimulus units in this analysis does not indicate the significance of stimulus units independent of differences in wave length, since the two are confounded.

Regarding the postulated equivalence of stimulus units of differing size Mednick and Freedman (1960) report that this equivalence tends to break down at the ends of the continuum being tested. Inspection of Fig. 2 reveals that over the range of unit size employed in this study (5–20 mμ.) there was no evidence to indicate that units of differing physical size were behaviorally equivalent. Thus, although the evidence gathered in this study supports the first hypothesis offered by Mednick and Freedman (1960), it fails to support the second one. In agreement with the present experiment, a study by Gewirtz, Jones, and Waerneryd (1956) also demonstrated that number of stimulus units taken alone is not the sole determiner of generalization. That study employed physical units (visual angle) which differed in size from one group to the next, and the gradients obtained were clearly not equivalent. In fact, in that experiment the group with the smallest size of stimulus unit produced the steepest gradient, a surprising finding which is explained by the fact that the region of the visual field to which S is most sensitive was employed with this group.

As a result of this study, we conclude that with human Ss neither the physical difference nor the number of stimulus units separating a generalization test stimulus from the original will adequately account for generalized response strength, but rather that both must be taken into account in order for an accurate prediction to be made. This study adds to the growing body of evidence which indicates that in human Ss stimulus generalization is not simply a function of the physical stimulus differences between the CS and the novel test stimulus. A more complete interpretation requires that various psychological determinants must be taken into account.

REFERENCES

GEWIRTZ, J. L., JONES, L. V., & WAERNERYD, K. E. Stimulus units and range of experienced stimuli as determinants of generalization-discrimination gradients. *J. exp. Psychol.*, 1956, **52**, 51–57.

GIBSON, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **34**, 237–253.

HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **25**, 361–372.

HUMPHREYS, L. G. Generalization as a function of method of reinforcement. *J. exp. Psychol.*, 1939, **25**, 361–372.

KALISH, H. I. The relationship between discriminability and generalization: A reevaluation. *J. exp. Psychol.*, 1958, **55**, 637–644.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

THOMAS, D. R., & JONES, C. G. Stimulus generalization as a function of the frame of reference. *J. exp. Psychol.*, 1962, **64**, 77–80.

THOMAS, D. R., & LANIER, W. G. A comparison of stimulus generalization tendencies to respond and not to respond. *Psychol. Rec.*, 1962, **12**, 61–65.

THOMAS, D. R., & MITCHELL, K. The role of instructions and stimulus categorizing in a measure of stimulus generalization. *J. exp. Anal. Behav.*, 1962, **5**, 375–381.

(Received January 27, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 63–69

EFFECTS OF TRANSFER AND PROBLEM STRUCTURE IN DISJUNCTIVE CONCEPT FORMATION[1]

HERBERT WELLS[2]

Yale University

3 studies dealt with the attainment of disjunctive concepts. The positive instances of disjunctive concepts contain either Values A or B, while conjunctive concepts are those whose positive instances contain both A and B. The former are typically more difficult to attain. The 1st experiment showed that training on disjunctive concept problems increases the proportion of Ss who offer a disjunctive concept when a conjunctive concept is also available. The 2nd experiment compared inclusive (A or B or both) and exclusive (A or B but not both) disjunctive problems. The latter were found to be more difficult in terms of the time to solution.

Several recent studies have investigated the influence of the "logical structure" of the concept on the process of concept formation. Concept formation studies have typically used conjunctive concepts, in which all positive instances contain a common combination of elements (e.g., red circles); and relational concepts, in which a certain relation (e.g., equal to) is common to the positive instances. Bruner, Goodnow, and Austin (1956) also used a disjunctive concept, one which states that an instance is positive if it has either of two elements (e.g., red figures or circles). They concluded that for their *S*s, disjunctive concepts were more difficult to attain than conjunctive concepts. Hunt and Hovland (1960) found that when conjunctive, relational, and disjunctive concepts were equally correct solutions to the same problem, very few *S*s offered disjunctive solutions. It is assumed that in such a choice situation the frequency of each type of solution is a measure of its relative difficulty.

The present research explored some of the factors involved in the difficulty of disjunctive concepts. In this discussion, following Hovland (1952), a "dimension" will denote a way in which the instances can vary (color; shape; etc.), and the "value" of a dimension denotes the state of that dimension for a given instance (red; square; etc.).

One explanation of the difficulty of attaining disjunctive concepts would be that each of the positive values of

---

[1] This research is based on a dissertation submitted to Yale University in partial fulfillment of the requirements for the PhD degree. The author was supported by a United States Public Health Service Predoctoral Research Fellowship.

The author wishes to express a special debt to the late Carl I. Hovland, who served as the original chairman of his advisory committee. His suggestions and critical comments contributed in large part to the general form of the study. The author is also indebted to the members of his committee, Donald W. Taylor (who served as chairman during the final stages of the work), Sidney J. Blatt, and William Kessen, for their generous help and guidance.

Support for the preparation of this report for publication was provided by a grant from the Ford Foundation to Donald W. Taylor for his continuing program of research on thinking.

The author wishes to thank Paul A. Lane, Richard Ehmer, and Marjorie Ehmer of the University of Bridgeport Psychology Department for their help in recruiting *S*s and in providing space for the experimental sessions.

[2] Now at the University of Washington.

such a concept is found in some, but not all, of the positive instances. In the case of a conjunctive concept, on the other hand, both positive values appear in all of the positive instances. Thus the proportion of positive instances which contain each positive value is different for the two types of concepts. But Bruner et al. (1956) and Hunt and Hovland (1960, 1961; see also Hovland, 1960) have suggested that prior experience in learning certain kinds of concepts may also be a factor in the greater difficulty of disjunctive concept learning, and in *S*s' lack of preference for disjunctive solutions. Training on disjunctive concept formation might then be expected to transfer to a choice problem for which either a conjunctive or a disjunctive concept is possible. The first experiment reported below was designed to test the prediction that such transfer would increase the proportion of *S*s who choose the disjunctive solution on a choice problem.

The disjunctive concepts described above and used in earlier studies have always been of the "inclusive" variety. That is, an instance is positive if it has either Value A or Value B or both. It is also possible to specify an "exclusive disjunctive" concept, in which positive instances are those which contain either Value A or Value B but not both. The second experiment reported below was designed to determine which of the two types of disjunctive concepts is more difficult to attain.

EXPERIMENT I

Method

Design.—There were two experimental groups, E2 and E4. The *S*s in these groups were given a series of training problems, for which the only correct solutions were disjunctive concepts, followed by a choice problem which could be solved by either a conjunctive or a disjunctive concept. The *S*s in Group E2 received two training problems, followed by a choice problem, then two more training problems and a final choice problem. Those in Group E4 received four training problems followed by a choice problem. Each *S* in Groups E2 and E4 was given a preliminary choice problem at the beginning of the experiment, which provided a way of insuring the comparability of the groups and of assessing the prior tendency to select conjunctive and disjunctive concepts.

To control for the effects of warm up and exposure to the stimuli, a control group was given a series of eight different choice problems. The order of problems was varied in a Latin square design. The same eight choice problems were also given to different *S*s in Groups E2 and E4. In each experimental group *S*s could then be compared with control group *S*s at the appropriate point in the sequence of problems.

Problems and stimuli.—On each choice problem there were 12 "learning instances," from which *S*s formed a concept. Six of these learning instances were positive examples of both the conjunctive and the disjunctive concept. The other 6 were negative instances; that is, they were not examples of either concept. After forming a conjunctive or a disjunctive concept on the basis of these learning instances, *S* was required to apply his concept to a set of 12 "test instances." Six of the test instances were positive instances of the conjunctive concept but negative instances of the disjunctive concept. In the other 6 the situation was reversed. The *S* was asked to identify the 12 test instances as positive or negative according to the concept he had formed on the learning instances. The test instances which he identified as positive provided a measure of the concept *S* had formed.

Disjunctive training problems could be solved only by a disjunctive concept. Again there were 6 positive and 6 negative learning instances. On disjunctive training problems the 12 learning instances were later used as test instances in the sorting procedure.

Instances were drawn in ink on 3 × 5 in. cards. On each card there were four numbered positions in a horizontal row. Each position represented a dimension. In each position was a figure, which represented the value of that dimension on that instance. The figures used were triangles, circles, and squares of approximately equal size. Thus there were four dimensions of three values each. Concepts always involved two different figures (values), each in a different position (dimension). For example, a conjunctive concept might be "Circle in Position 1 and triangle in Position 3." On choice problems the conjunctive concept involved two of the positions and the disjunctive concept involved the other two. Positive instances were always called "Alpha cards" and negative instances were called "Not-Alpha cards." Control and experiment group *S*s were matched in terms of the dimensions which were relevant on the various problems.

Subjects.—The *S*s were 72 college freshmen, 26 men and 46 women, recruited from introductory psychology classes at the University of Bridgeport. They were randomly assigned to experimental conditions.

Procedure.—The *S*s were seen individually in 45-min. sessions. Each *S* was seated at a small table across from the *E*. He was told that he would be asked to find the difference between Alpha and Not-Alpha cards. He was instructed to state a rule or a set of rules which could later be used to sort out a shuffled set of cards into Alpha and Not-Alpha categories. The dimensions and values of the cards were explained.

All 12 cards were then placed on the table, Alpha cards on *S*'s right, and Not-Alphas on the left. The *S* could rearrange the cards if he wished. All *S*s were allowed 10 min. on each problem. After *S* had stated his rule, all the cards were taken up and *S* was given a shuffled deck of 12 cards to sort out by applying his rule.

The first choice problem functioned in part as a screening device, since pretests had indicated that few *S*s who failed to solve it would finish in the time available. Sixteen *S*s were dropped for failure to find any adequate rule on the first problem. In addition, 7 *S*s assigned to the two experimental groups who gave adequate but nonconjunctive rules on the first problem were dropped, after it had been found that all *S*s in the control group gave conjunctive rules. Seven other *S*s were dropped from the experimental groups for failure to solve two consecutive disjunctive training problems, since it was not administratively feasible to keep them longer. (Any selective bias introduced by dropping from the experimental groups those *S*s who could not attain disjunctive concepts would appear to be balanced by having dropped those who did not give conjunctive solutions on the first problem. But in addition, the major analyses of the data were repeated assuming the inclusion of these 7 *S*s in the experimental groups, and assuming on their part conjunctive solutions to the critical choice problems. Essentially the same results were found.)

If *S* sorted according to the conjunctive concept on the first problem *E* said "Okay, that's the idea." The *S*s in the control group were told that on subsequent problems *E* would not tell them whether or not they were correct. In the two experimental groups, on the other hand, the *S*s were told after every correct disjunctive sorting, "Okay, that's right."

Results

The principal comparison is that of the experimental groups on the critical (second) choice problem with the control group on a corresponding choice problem. The major dependent variable was derived from *S*s' sorting of the test instances. The data may be analyzed in terms of the number of *S*s in each group who sort the test instances in a purely conjunctive, purely disjunctive, or mixed fashion. Table 1 presents a chi square analysis of these frequency data. The results of the first, second, and fourth comparisons should be interpreted somewhat conservatively in view of the small expected frequencies in some of the cells. However, it is clear that both Groups E2 and E4 differ from

TABLE 1

NUMBER OF *S*S SORTING CORRECTLY BY EACH CONCEPT

Group	Conjunctive	Disjunctive	Mixed	χ^2
Control	23	0	1	12.874**
E2 (second choice problem)	12	4	8	
Control	24	0	0	14.270***
E4	13	7	4	
E2 (second choice problem)	12	4	8	2.192
E4	13	7	4	
Control	24	0	0	12.632**
E2 (third choice problem)	14	4	6	

** $p < .01$.
*** $p < .001$.

the control group, but that there is not a significant difference between E2 and E4.

The most obvious effect of disjunctive training was to produce a shift away from conjunctive solutions. This effect may be tested by comparing the frequency of conjunctive solutions with the combined frequencies of disjunctive and mixed solutions for each group. For Group E2 vs. the control group, $\chi^2_c(1) = 10.549$; $p < .005$. For Group E4 vs. the control group, $\chi^2_c(1) = 11.794$; $p < .001$. However, the two experimental groups do not differ from each other (χ^2_c = zero).

The major prediction may be tested by comparing the frequency of disjunctive solutions with the combined frequencies of conjunctive and mixed solutions. On such a comparison, Group E4 differs from the control group, $\chi^2_c(1) = 6.021$; $p < .02$, but Group E2 does not, $\chi^2_c(1) = 2.455$; $p > .10$.

It appears that disjunctive training also results in an increase in mixed solutions. If the frequency of mixed solutions is compared with the combined frequencies of conjunctive and disjunctive solutions, Group E2 differs from the control group, $\chi^2_c(1) = 4.923$; $p < .05$, but Group E4 does not, $\chi^2_c(1) = 2.455$; $p > .10$. Again the two experimental groups do not differ from one another, $\chi^2_c(1) = 1.000$; $p > .25$.

A somewhat more sensitive analysis is also possible. Six of the test instances could be correctly identified as Alpha by the conjunctive concept, the remaining six by the disjunctive. A composite score was assigned to each *S*, based on the difference between the number of disjunctive and the number of conjunctive test instances which he identified as Alpha. The composite score represents an index of "disjunctivity." The scores of *S*s in each experimental group were separately compared with those of the control group at a corresponding point in the sequence of problems. Groups E2 and E4 were compared on their respective critical problems, which were given at different stages of training. Since the distributions of scores were distinctly nonnormal, ranked data were analyzed by the Mann-Whitney *U* test, corrected for the number of tied ranks (Siegel, 1956, pp. 116 ff.). This analysis indicated that both experimental groups were more disjunctive than the control group. (For Group E2 vs. the control group, $U = 426$; $z = 3.639$; $p < .0002$. For Group E4 vs. the control group, $U = 420$; $z = 3.709$; $p < .0002$.) However, the two experimental groups do not differ from each other ($U = 301.5$; $z = 0.303$; $p > .38$). The distributions of scores suggest that any difference between Groups E2 and E4 must be concentrated in the scores of those *S*s who do not sort the cards on a purely conjunctive basis. A comparison of the "nonconjunctive" *S*s in the two experimental groups shows that there is a nonsignificant tendency for those in Group E4 to be more disjunctive in their sorting than those in Group E2, $U(11,12) = 42.5$; critical U for $p = .05$ is 38.0. The control group may be used as a baseline by dropping any experimental group *S*s who have the same score as a control group *S* (1 *S* in Group E2). The results are similar ($U = 37.0$; critical $U = 34.0$). It should be pointed out that either estimate of the difference is probably somewhat conservative. Since the number of cases is too small to convert *U* to *z*, it is impossible to correct for the number of tied ranks.

It is also of interest to examine the verbal rules or concepts offered by

*S*s. Two judges independently scored the concepts with 100% agreement. Table 2 presents the data on verbal rules. If the frequency of disjunctive rules is compared with the combined frequency of conjunctive and mixed rules, both experimental groups differ from the control group. For the comparison of Group E2 with the control group, $\chi^2_c(1) = 10.232$; $p < .01$. For the comparison of Group E4 with the control group, $\chi^2_c(1) = 11.794$; $p < .001$. Groups E2 and E4 did not differ from one another. Inspection of *S*s' rules indicates that those whose rules were scored disjunctive but who sorted the cards in a mixed fashion mentioned more than two values in their rules (e.g., "Alphas have a circle in Position 2 or squares in 3 and 4").

Comparisons of the third choice problem of Group E2 with the corresponding choice problem of the control group (seventh problem) are presented in Tables 1 (sorting) and 2 (rules). There is practically no difference between the second and third choice problems for the E2 *S*s in the number of disjunctive and mixed choices. In terms of ranked sorting scores for all *S*s there is no significant difference between the third choice problem in E2 and the second choice problem in E4, both of which were given after four disjunctive training problems ($U = 326.0$, $z = 0.871$, $p > .18$).

TABLE 2

NUMBER OF *S*s GIVING RULES JUDGED CONJUNCTIVE, DISJUNCTIVE, MIXED

Group	Conjunctive	Disjunctive	Mixed	χ^2
Control	23	0	1	13.790**
E2 (second choice problem)	12	10	2	
Control	24	0	0	14.270***
E4	13	11	0	
E2 (second choice problem)	12	10	2	2.088
E4	13	11	0	
Control	24	0	0	12.632**
E2 (third choice problem)	14	7	3	

** $p < .01$.
*** $p < .001$.

EXPERIMENT II

Method

Problems and stimuli.—Three inclusive disjunctive (D1) and three exclusive disjunctive (D2) problems were used. Each D1 problem was matched with a D2 problem in terms of relevant dimensions and positive values. The stimuli were the same as the distributed stimuli used in Exp. I and II.

Design.—Each *S* solved one D1 problem and one D2 problem. The two types of problems were given in counterbalanced order to different *S*s. The first and second problems given to any *S* always had one relevant dimension in common. The problems were given in sequences such that no matter which problem (D1 or D2) came first, the same change in relevant dimensions occurred between the first and second problems. For example, if the relevant dimensions on the first problem were Positions 1 and 2, those on the second problem were 1 and 3; this was the case whether a D1 or a D2 problem came first. Six different sequences of problems were used.

Subjects.—Twelve Yale undergraduates, recruited from first-yr. psychology classes, served as *S*s. The *S*s were randomly assigned to the six sequences of problems, 2 *S*s to each sequence. They were seen in individual sessions lasting from 20 to 40 min. No *S*s were dropped.

Procedure.—The procedure was essentially the same as that in Exp. I, as were the instructions to *S*s. Time-to-solution scores were taken on each problem, and were used as the dependent measure.

In several cases *S*s working on D2 problems took an inordinate amount of time without stating any concept, or none which did not reduce to memorization. If *S* had worked on a problem for 15 min. without having stated a concept or located at least one positive value, *E* rearranged the Alpha cards so that the three containing one positive value were grouped together, and the other three formed another grouping. For none of the 4 *S*s thus instructed did this procedure lead to an immediate solution; they took an additional 102 to 499 sec. before offering an adequate rule.

Results

Since all *S*s worked both kinds of problems, in counterbalanced order, the principal dependent measure is the difference between *S*'s time on the D2 problem and his time on the D1 problem. In all 12 cases the difference (D2 minus D1) was positive. It is therefore concluded that the D2 problems were more difficult ($p < .001$, sign test). Moreover, there was no overlap between times on the two kinds of problems.

It is also possible to evaluate transfer or order effects. The median time scores in sec. on D1 and D2 problems, respectively, were 202.5 and 1111.5 for those *S*s who received a D2 problem first, but 143.0 and 770.5 for those who received a D1 problem first. It would appear that solution of D2 problems was facilitated when they occurred last, but that this was not so for D1 problems. This effect is also reflected in the median difference scores for the two problem-order groups: 1042.0 for those who received a D2 problem first but 528.0 for those who received a D1 problem first. The Mann-Whitney *U* test was applied to the ranked difference scores for the two (problem-order) groups, and indicated that the differential transfer effect was almost significant, $U(6) = 6.0$; $p = .064$. This may reflect in part a tendency for more facilitation to result from an easy-to-hard order.

DISCUSSION

The results of Exp. I indicated that a moderate amount of training on disjunctive concept problems increases the probability that *S*s will offer the disjunctive solution to a problem for which a conjunctive solution is also possible. It appears unlikely that the preference for conjunctive concepts is solely a function of the relative proportion of positive instances which contain the positive values of each concept.

Previous (extraexperimental) learning may account in part for the initial difficulty of disjunctive concept problems. On the disjunctive training problems *S*s apparently learned something about the methods of solution appropriate for disjunctive concepts. These methods may be incompatible to some extent with those which are appropriate for conjunctive solutions. A number of *S*s remarked that the disjunctive training problems had forced them to adopt the strategy of searching a single dimension at a time across all the cards. The conjunctive solution, on the other hand, demanded that *S*s notice the joint occurrence on each card of a pair of values on different dimensions. Several *S*s developed the habit of scanning the Not-Alpha cards to determine which values were missing. They would then state that no Not-Alpha card contained either Value A or Value B. This strategy allows *S* to escape noticing the more frequent occurrence of the conjunctive values in the Alpha cards. It has been found that *S*s are initially less inclined to focus on negative instances than on positive instances (Bregman & Hovland, 1961), and that the information conveyed by negative instances is less efficiently utilized (Hovland & Weiss, 1953). Bruner et al. (1956) have suggested that efficient solution of disjunctive problems depends upon the use of negative instances. It may be that in some cases training functioned to modify prior habits of focusing exclusively on the positive instances.

The most striking effect of disjunctive training was a shift away from conjunctive concepts. What appeared to distinguish the two experimental groups on their respective critical choice problems was the increase from E2 to E4 in the efficiency with which *S*s formulated the disjunctive solution. However, with the amount of training used in this study, this was not a significant difference.

The behavior of *S*s in Group E2 on the third choice problem deserves comment. The amount of training preceding this problem was the same as that preceding the second choice problem in Group E4. But the results on this problem more nearly resemble those on the second choice problem in Group E2, although none of the differences was significant. It is possible that the second choice problem, for those who offered a conjunctive solution to it, broke the disjunctive "set" established on the first two training problems.

An objection which might be raised against Exp. I is that *S*s in the experimental groups were merely learning that *E* expected them to offer a disjunctive solution. They might, therefore, have found both solutions but have given only the disjunctive. Two observations suggest that this was unlikely. Several *S*s in the experimental groups who found the conjunctive solution on a critical choice problem expressed surprise and pleasure as they stated the concept. The *S*s who gave the disjunctive solution to their final problem were again shown the array of learning cards, and were asked if they could find any other answer. None of them stated the conjunctive concept immediately, but when they had found it they too commented that the conjunctive concept was a "better" or "more natural" solution.

The results of Exp. II indicated that *S*s take more time to find the solution to exclusive disjunctive (D2) problems than to inclusive disjunctive (D1) problems. On D2 problems, *S*s appeared to have most difficulty accounting for the presence on the Not-Alpha side of cards containing both positive values.

REFERENCES

BREGMAN, A. S., & HOVLAND, C. I. The effect of the labelling of instances on concept learning. Unpublished manuscript, Yale University, 1961.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

HOVLAND, C. I. A "communication analysis" of concept learning. *Psychol. Rev.*, 1952, **59,** 461–472.

HOVLAND, C. I. Computer simulation of thinking. *Amer. Psychologist,* 1960, **15,** 687–693.

HOVLAND, C. I., & WEISS, W. Transmission of information concerning concepts through positive and negative instances. *J. exp. Psychol.*, 1953, **45,** 175–182.

HUNT, E. B., & HOVLAND, C. I. Order of consideration of different types of concepts. *J. exp. Psychol.*, 1960, **41,** 23–29.

HUNT, E. B., & HOVLAND, C. I. Programming a model of human concept formulation. *Proc. West. Joint Comput. Conf.*, 1961, 145–155.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received January 27, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 70–74

EFFECT OF INTERTRIAL REINFORCEMENT ON THE AFTEREFFECT OF NONREINFORCEMENT AND RESISTANCE TO EXTINCTION

E. J. CAPALDI, DICK HART, AND LARRY R. STANLEY

University of Texas

2 groups of rats were trained to traverse a straight alley under identical patterns of partial reinforcement; a 3rd group received consistent reinforcement. Both partial groups received intertrial reinforcements, one following nonreinforced trials preceding reinforced ones, the 2nd following reinforced trials preceding either reinforced or nonreinforced ones. The 1st partial group was not more resistant than the consistent group, while the 2nd partial group showed the typical partial reinforcement effect.

One of the most direct and powerful means available for testing the several propositions which comprise Hull's (1952) aftereffect hypothesis is to prevent the aftereffect of nonreinforcement from becoming conditioned to the instrumental response. Sheffield (1949) attempted this by utilizing a spaced-trial procedure. A similar attempt was made here by employing an experimental procedure termed intertrial reinforcement. The intertrial reinforcement procedure and the experimental design are described below.

The intertrial interval which *S* usually spent in a neutral waiting box was sometimes divided into two equal intervals, only the initial half of which was spent in the waiting box. The remaining half of the interval was spent by *S* in the goal box where it was placed directly; i.e., *S* did not run into the goal box but, rather, was placed there by *E*. On such placement or intertrial reinforcement trials the goal box was always baited with food. For one partial group (PN) the intertrial reinforcements were given following nonreinforced runway trials which preceded reinforced ones, and never following any other trial. The intertrial reinforcements of the second partial group (PR) were given following reinforced runway trials which preceded either reinforced or nonreinforced runway trials. A consistently reinforced group (C) received the same number of intertrial reinforcements as did each of the partial groups.

The relevance of the present experiment to Hull's hypothesis can now be indicated. Hull assumed that nonreinforced trials occasion distinctive aftereffects. These stimuli may persist until the following trial, and they become conditioned to the instrumental response only on reinforced trials. Under partial reinforcement the aftereffects of nonreinforcement are conditioned to the locomotor response on reinforced trials preceded by nonreinforced ones. In extinction only the aftereffects of nonreinforcement occur. Since these stimuli have been directly conditioned to the instrumental response under partial training but not under consistent, resistance is expected to be greater in the partial group.

If intertrial reinforcement modifies or interferes with the aftereffect of nonreinforcement, substituting for it the aftereffect of reinforcement, it can be expected that Group PN will not be more resistant to extinction than Group C because Group PN's intertrial reinforcements are given following all nonreinforced trials which precede reinforced ones. Thus, stimuli characteristic of nonreinforcement do not have the opportunity to become conditioned to the instrumental response in Group PN. Group PR, which is a control for the effects of intertrial reinforcement, should, of course, show the typical partial reinforcement effect of increased resistance to extinction. In this group, the intertrial reinforcements are not given following nonreinforced trials and, therefore, stimuli characteristic of extinction should be directly conditioned to the instrumental response just as they are in the typical partial reinforcement situation not employing intertrial reinforcement.

METHOD

Subjects.—The *S*s were 21 female and 15 male experimentally naive albino rats from the colony maintained by the Psychology Department, the University of Texas. The *S*s were approximately 100 days old at the beginning of the experiment.

Apparatus.—The apparatus was a straight-alley runway with an overall length of 8 ft. and width of 4 in., enclosed by 9-in. high sides. It was constructed of wood and painted gray. The goal area, covered with hinged, ½-in. hardware cloth, was separated from the remainder of the runway by a painted, sheet metal guillotine door located 1 ft. from the extreme distal end of the alley. Mounted side by side on runners to the right and extreme distal end of the goal area were two identical-appearing, 3 in. wide × 4¼ in. long compartments which could be moved into position noiselessly. On a particular trial *S* had access to only one of these compartments, the other being both inaccessible and hidden from *S*'s view. One compartment was baited with a wet mash and used on rewarded trials. The remaining unbaited compartment was used on nonrewarded trials. The alley portion of the runway was covered with hinged, $\frac{1}{16}$-in. thick Plexiglas. Running speed was measured by a Standard Electric timer.

Placing *S* on a treadle resting over a microswitch in the floor of the alley and positioned at the extreme proximal end of the alley started the clock. The clock was stopped when *S* interrupted a photobeam located 4 in. from the extreme distal end of the runway. It can be seen that the compartments described above were so placed and the position of the photocell was such that *S* could not determine whether or not the trial was to be reinforced prior to the time the photocell beam was broken. The distance from the tip of the treadle to the light beam was 7 ft. The entire apparatus was painted gray. A box 11½ in. square and 12 in. deep, resting on 20-in. high legs, and painted entirely black served to contain *S* between trials.

Preliminary training.—On Days 1–3 *S*s were housed in large cages in groups of 4 or 5. On each of these days, as on each succeeding day of the experiment, *S*s were fed a vitamin-supplement diet of wet mash for 1 hr. in the home cage. On Day 4 *S*s were placed in smaller individual cages. On Days 4–9 *S*s were handled on a large table for 1 hr. per day in groups of 6 or 7. The groups were handled successively rather than concurrently. On each of these days, wet mash was available on the table and *S* was fed in the home cage immediately following the termination of the handling period. On Day 10 *S*s were handled as usual; however, feeding time on the table was reduced by 30 min. On Day 11 each *S* explored the unbaited runway alone for 5 min. Following the exploratory training, *S* was fed for 10 min. on the table and then returned to the home cage where it was immediately fed for 1 hr. On Day 12 each *S* was hand-placed in the baited goal box of the runway. Each placement lasted 30 sec. with 30 sec. spent in the neutral waiting box elapsing between placements. Since this training was given to insure that *S* would eat during the intertrial reinforcements of the experimental phase, a training criterion was adopted. Each *S* was to be given a minimum of five placements and training would continue until *S* had eaten in at least three placements. Only 1 *S* required more than five placements to reach the criterion; that *S* required seven placements.

Experimental training.—At the end of Day 12, 7 male and 5 female *S*s were assigned to each of the three groups of 12 *S*s each. One *S* from the C group died during the course of the experiment. The data for this *S* were discarded. Each group received 3 days of acquisition training and 2 days of extinction training, 10 trials per day. A trial was begun by placing *S* on the treadle. When

TABLE 1

Reinforcement Schedule and Intertrial Reinforcements of the Partial Groups

Day	Trial 1	2	3	4	5	6	7	8	9	10
1	R	R	N	N	N +	R ×	N	N +	R ×	R
2	N +	R	R	R ×	N	N	N +	R ×	R	N
3	R ×	R	N	N +	R ×	N +	R ×	N	N +	R

S interrupted the photobeam, the guillotine door was lowered. The *S* was confined to the goal box for 15 sec. following each run. The intertrial interval, spent in the waiting box, was 30 sec. On intertrial reinforcement trials, *S* spent only the initial 15 sec. of the 30-sec. intertrial interval in the neutral box. The remaining 15 sec. was spent in the baited goal box where *S* was placed by *E*. At the termination of the intertrial reinforcement trial, *S* was immediately brought forward to the start portion of the runway, and the next trial was begun.

Group C was rewarded with a wet mash following each run. The 50% schedule under which Groups PR and PN were trained is shown in Table 1, the letter R representing a reinforced trial, the letter N, a nonreinforced one. The symbol + indicates the position in the series at which Group PN's intertrial reinforcements were given, the ×, those of Group PR. Notice that both groups received the same daily and total number of intertrial reinforcements and over days the ordinal or trial position of the intertrial reinforcements was the same for both groups, with Group PN's intertrial reinforcement occurring following nonreinforced trials preceding reinforced ones and Group PR's following reinforced trials which preceded either reinforced or nonreinforced ones. One-half the *S*s of Group C received intertrial reinforcements following the same trials each day as Group PN and the remaining half following the same trials each day as Group PR.

Following the 3 days of acquisition training, extinction training, which consisted of 10 nonreinforced trials per day on each of 2 consecutive days, was begun. If an *S* failed to enter the goal box within 120 sec., it was removed from the alley for the 30-sec. intertrial interval, assigned a time score of 120 sec., and then given the next trial. The goal-box confinement period remained at 15 sec. in extinction.

Results

A Kruskal-Wallis (Siegel, 1956) one-way analysis of variance on the sum of the running times on the 10 trials of the last day of acquisition indicated that the groups did not differ reliably, $\chi^2(2) = .79$. The mean time in sec. on the last 10 trials was for Group C, 9.13, for Group PN, 9.09, and for Group PR, 8.98. The mean number of trials required by each group to run as slow or slower than 25 sec., 50 sec., or 75 sec. in extinction are presented in Table 2. In the event that an *S* failed to meet a particular criterion, a trial score of 21 was assigned. Since it is reasonable to suppose that a trial score of 21 underestimates the true mean, statistical comparisons between Group PR and the remaining groups at the 50-sec. criterion and between Groups C and PN and Groups PR and PN at the 75-sec. criterion probably underestimate the actual magnitude of the differences. Since the means of both Group C and PR may be underestimated at the 75-sec. criterion, these groups were not compared statistically at this point.

Analyses of variance on the means of the various criteria indicated that the groups differed significantly at the 25-sec. criterion, $F(2,32) = 5.72$, $p < .01$; at the 50-sec. criterion, $F = 5.50$, $p < .01$; and at the 75-

TABLE 2

Mean Number of Trials Required to Run as Slow or Slower in Extinction Than the Indicated Criterion and Percentage of *S*s Meeting the Criterion

Group	Criterion 25 Sec. Trials	25 Sec. %	50 Sec. Trials	50 Sec. %	75 Sec. Trials	75 Sec. %
C	4.91	100	7.45	100	14.09	64
PN	7.25	100	8.50	100	10.08	100
PR	9.42	100	14.33	75	17.92	42

sec. criterion, $F = 5.27, p < .02$. Application of an extension of Duncan's multiple range test to group means with unequal numbers of replications (Kramer, 1956) to the individual means at the various criteria indicated that at the 25-sec. criterion only the difference between Groups PR and C was significant ($p < .05$), while at the 50-sec. criterion Group PR differed significantly from Group C ($p < .01$) and from Group PN ($p < .05$), and at the 75-sec. criterion only the difference between Group PR and PN ($p < .01$) was reliable. Thus, Group PN failed to manifest greater resistance than Group C at any of the criteria, while Group PR showed the typical partial reinforcement effect of increased resistance to extinction at the 25-sec. and 50-sec. criteria.

Figure 1 presents the mean running times in extinction for Day 1 and Day 2 in blocks of two trials. On the whole, Group PN ran slightly more rapidly than Group C in the early extinction trials, but later in extinction Group C ran faster than Group PN. Group PR ran faster than Group PN throughout extinction and, with the exception of the final block of trials, faster than Group C throughout extinction. A Kruskal-Wallis test employing the sum of the Day 1 running times yielded $\chi^2(2) = 6.39$, which is significant beyond the .05 level. Subsequent Mann-Whitney U tests indicated that Group PR differed from Group C well beyond the .05 level, $U(11,12) = 28$, and from Group PN at the .05 level, $U(12,12) = 37$. The difference between Group PN and C was not significant.

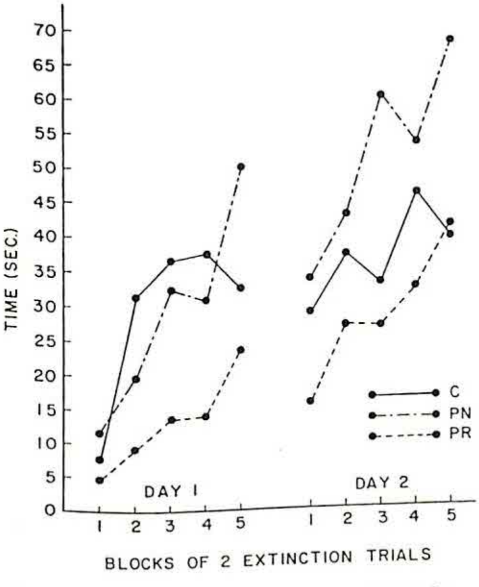


FIG. 1. Running speeds during extinction.

The Kruskal-Wallis test performed employing the Day 2 running times yielded $\chi^2(2) = 4.77$, which is insignificant ($p < .10$). Subsequent U tests indicated that Groups PR and PN differed at the .05 level, $U(12,12) = 37$. Neither of the two remaining differences was significant.

DISCUSSION

The major results of the present experiment seem to be these. Group PN was not more resistant to extinction than Group C, while Group PR was significantly more resistant to extinction than either Group C or PN. The relatively rapid extinction of Group PN is ascribed to a failure of the aftereffect of nonreinforcement to become conditioned to the locomotor response due to the interfering properties of intertrial reinforcement and, therefore, these results are taken to support the Hull-Sheffield hypothesis.

The present results might be considered as quite unexpected in view of an experiment by Katz (1957). In that experiment, somewhat similar in design and conception to the present one, results were obtained which are in apparent disagreement with those reported here. Katz trained rats so that trials in one runway alternated with trials in a second runway. One group of rats was given partial reinforcement in the first runway and consistent in the second (P-C), a second group was given consistent reinforcement in the first runway and partial in the second (C-P), while a third group was consistently reinforced in both runways (C-C). All groups were given extinction training in the second runway. Group C-P was most resistant to extinction, followed by Group P-C and, finally, Group C-C was least resistant to extinction. In the Katz experiment the reinforcement designed to interfere with the aftereffect of nonreinforcement (Group C-P) was administered in a goal box which differed from the one in which the aftereffect of nonreinforcement was originally occasioned in terms of length, color, and other physical properties. In the present experiment the intertrial reinforcement was administered in the same goal box in which the aftereffect of nonreinforcement was occasioned. It would appear, therefore, that the effectiveness of an intervening reinforcement in modifying the aftereffect of nonreinforcement is a function of the stimulus similarity which obtains between the original reinforcement situation and the interpolated or interfering situation.

The intertrial reinforcement procedure with respect to Group PN is similar to the typical delay of reinforcement procedure. These procedures are similar, of course, in that under both *S* initially enters an empty, i.e., nonbaited, goal box, and following a waiting period is there reinforced without having again to execute the instrumental response. However, under the delay procedure *S* spends the entire waiting period in the goal box, while under the intertrial reinforcement procedure only a portion of the waiting period is spent in the goal box, the remainder of the wait occurring in the intertrial interval box. That the present results can be understood in terms of the similarity between the two procedures seems unlikely, however. Delayed reinforcement on some portion of the acquisition trials (Crum, Brown, & Bitterman, 1951) as well as a combination of delayed and partial reinforcement (Peterson, 1956) give rise to greater resistance to extinction as compared with consistent reinforcement. Group PN, of course, was not more resistant than Group C.

REFERENCES

CRUM, J., BROWN, W. L., & BITTERMAN, M. E. The effect of partial and delayed reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1951, **64**, 228–237.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KATZ, S. Stimulus aftereffects and the partial reinforcement extinction effect. *J. exp. Psychol.*, 1957, **53**, 167–172.

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 1956, **12**, 307–310.

PETERSON, L. P. Variable delayed reinforcement. *J. comp. physiol. Psychol.*, 1956, **49**, 232–234.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 411–426.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

(Received January 31, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 75–79

EFFECT OF VERBAL, VISUAL, AND AUDITORY AUGMENTING CUES ON LEARNING A COMPLEX MOTOR SKILL[1]

LAWRENCE KARLIN AND RUDOLF G. MORTIMER[2]

New York University

In comparing the effects of various additional cues on performance of a compensatory tracking task, a verbal cue group was significantly superior in final training and transfer trials to a control group. Both visual and auditory cues were superior to the control but individually none of these differences was significant. Comparison of these with earlier results suggested that augmenting cues were most effective in transfer when they functioned to reward desirable patterns of behavior and develop standards of performance rather than to guide immediate action.

It was the aim of the present study to compare the effects of adding different types of cue to the primary display in learning compensatory tracking tasks, and to evaluate their effectiveness for training purposes on the basis of the resultant performance during transfer tests when these cues were removed.

While a number of studies (Bilodeau, 1952; Goldstein & Rittenhouse, 1954; Houston, 1947; Morin & Gagné, 1951) have shown that augmenting cues which are coincident or synchronized with the primary cue and more convenient to use than the primary cue materially improve performance on training tests, this improvement is lost in transfer tests.

Noncoincident types of cue in which information concerning performance is provided at the end of a trial or delayed in some other way may be effective in transfer tests. This has been shown by Goldstein and Rittenhouse (1954), Reynolds and Adams (1953), and Smode (1958). However, it should be noted that the results of Reynolds and Adams were not confirmed by Archer, Kent, and Mote (1956) and Archer and Namikas (1958).

It is apparent, then, that the differential effects of coincident and noncoincident augmenting cues on transfer trials are not clear-cut. The present study investigates this problem further using a verbal delayed feedback cue, and visual and auditory coincident type cues.

METHOD

Subjects and Apparatus

The *S*s were 60, paid, volunteer, right-handed male college students.

The apparatus consisted of three major parts: (*a*) *S*'s control and display assembly, (*b*) an electronic gating system and recording unit, and (*c*) a tracking course generator and timing control.

The *S*'s control consisted of a potentiometer controlled by a smooth round knob $1\frac{1}{4}$ in. in diameter set in a small console. The display consisted of a Dumont 304-A cathode ray oscilloscope whose tube face was covered by an orange filter on which were inscribed thin vertical lines $\frac{13}{64}$ in. apart. The center line which represented the target reticle was somewhat deeper and wider than the other lines so that when the CRT face was illuminated from the rear, the center line was seen as a markedly brighter white line than the other lines. Apart from a central horizontal slit $1\frac{5}{16}$ in. wide, the tube face was blanked off and was surrounded by a large matte gray panel. A Dumont 5-ADP-11 short persistence CRT was used in the oscilloscope.

---

[1] This research was part of a program carried out under contract with the United States Naval Training Device Center, Port Washington, New York, and is described in Technical Report No. 558-2.

[2] Now at Purdue University.

The cursor consisted of a blip on the tube face $\frac{3}{64}$ in. in diameter. The control-display relationship was such that 32° of control knob movement produced 1 in. of cursor displacement.

The visual augmenting cue was provided by placing two vertical red lines $\frac{3}{64}$ in. thick a distance of ±.25 in. on either side of the target reticle.

The electronic gating system and the recording unit consisted of 15 channels in which high speed counters cumulated the time that tracking performance lay within the range of each channel. From two of these channels time-on-target information was obtained. Two additional gating circuits controlled the auditory cue. The latter consisted of an 800-cps tone fed into a headset and sounded whenever *S*'s error was within ±.25 in. from the target line. For a more detailed description of the gating system, the reader is referred to Karlin (1961).

The tracking course generator was designed to control the duration of a trial (36 sec.), the interval during which scoring took place (the last 30 sec.), the appearance of the cursor at the beginning of a trial and its removal at the end of a trial, and the presentation of the tracking input signal. The latter consisted of a simple sine wave having a frequency of 10 cpm with a second and third harmonic of half the amplitude of the fundamental superimposed upon it; and this forced the cursor in a horizontal path along the CRT face. This section of the apparatus is described in greater detail by Karlin and Mortimer (1961).

Procedure

Fifteen *S*s were randomly assigned to each of the three experimental conditions and to the one control condition of the experiment. The *S* was seated before the display and operated the control with his right hand. His task was one of compensatory tracking such that this response was to counteract the horizontal movements of the cursor and maintain it on the target line.

Each *S* was scheduled for 5 consecutive days. On Days 1–3, *S*s in the experimental groups received 15 learning trials with augmenting cues, and on Day 4 they received 20 transfer trials during which the augmenting cue was not presented. The *S*s in the control group received the same trial sequence but augmenting cues were not provided at any time. Although data were collected for 4 days, *S*s were scheduled for 5 days in order to minimize possible end-spurt effects.

A trial began when the cursor appeared on the screen of the oscilloscope and ended when it disappeared from the screen. Each trial lasted a total of 36 sec., the last 30 sec. only being scored. An intertrial interval of 30 sec. was used.

Instructions were read to each *S* prior to data collection. The *S*s in the visual augmenting feedback group were instructed to the effect that the area between the two red lines defined a zone of minimally acceptable performance but that the *S*'s aim should be to keep the cursor on the target reticle. The *S*s in the auditory group received similar instructions with respect to the auditory cue. The *S*s in the verbal cue condition were given a score at the end of each learning trial which they were told could range from 0–1500 and which was a direct function of the time for which they had maintained the cursor within some minimally acceptable error zone about the target reticle. On Day 4, the experimental groups were told that the augmenting cues would not be present.

At the beginning of each trial the control knob was at its maximum counterclockwise position. All *S*s wore headphones throughout the experiment. The experimental room was dimly illuminated. A fan was used to provide masking noise. The oscilloscope was given a minimum warm-up time of 1 hr. and the cursor was checked for alignment prior to data collection of any *S*.

RESULTS

For each *S* on each trial, time within scoring tolerance (sec.) was obtained. This is defined as the time within an error of ±.25 in. about the target reticle. Integrated error scores were also obtained but since they yielded quite similar results only the former score is reported.

Figure 1 shows that the groups differed consistently from each other on learning and transfer trials with the verbal being superior to the visual, auditory, and control groups, in that order. A simple variance analysis was made of the data plotted in Fig. 1.

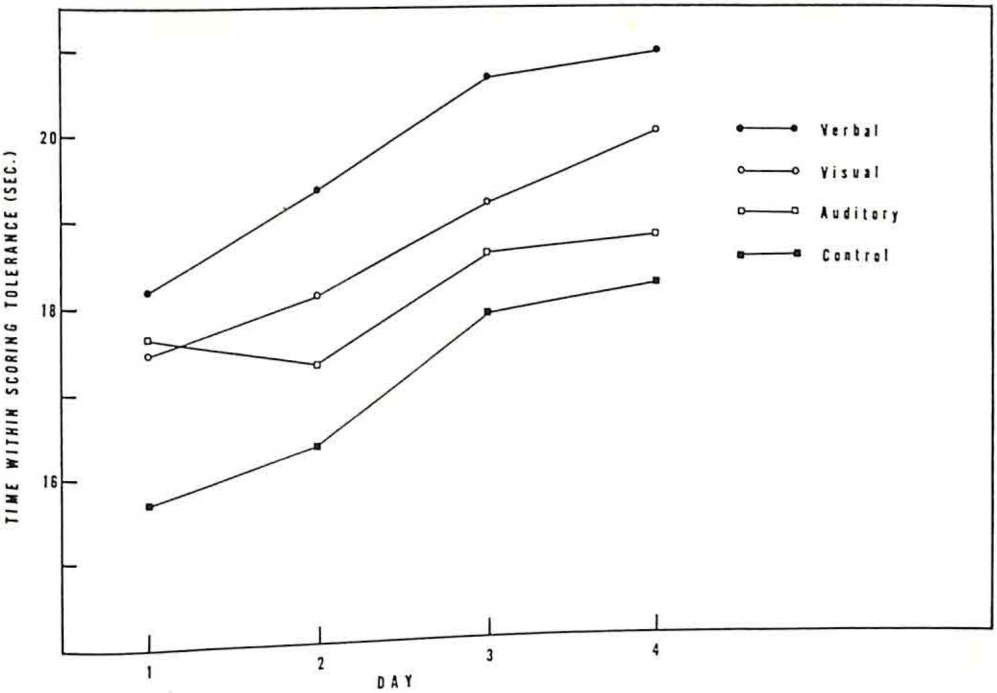


FIG. 1. Mean time within tolerance by day for each group. (Each point was obtained from the mean of the last five trials for each S and averaged over 15 Ss.)

The first and second days yielded nonsignificant Fs of 1.79 and 2.72 while the third and fourth days yielded Fs of 3.88 and 2.95 which were significant at the .05 level. The within-groups MSs, each based on 56 df, were used as error estimates for the above and were, respectively, (in squared sec.) 9.58, 8.97, 5.41, 7.88. These values show that individual differences diminished over the 3 training days and tended to rise somewhat on the fourth or test day.

TABLE 1

DUNCAN RANGE TEST OF MEAN DIFFERENCE BETWEEN CUES ON TIME WITHIN TOLERANCE IN SEC. FOR DAYS 3 AND 4

Comparison	Mean Difference	
	Day 3	Day 4
Verbal-Visual	1.48	.98
Verbal-Auditory	2.08*	2.22*
Verbal-Control	2.78**	2.60**
Visual-Auditory	.60	1.24
Visual-Control	1.30	1.80
Auditory-Control	.70	.56

* $p = .05$.
** $p = .01$.

Comparisons between individual means made by means of a Duncan range test (Edwards, 1960) are shown for Days 3 and 4 in Table 1. Note that the verbal group was significantly superior to the auditory and control groups in the last day of training and maintained this superiority in the transfer trials.

DISCUSSION

The results show that the noncoincident verbal cue was more effective than the coincident visual and auditory cues in all trials. It is possible that the verbal cue in not specifying a fixed tolerance limit for acceptable performance may have encouraged S to set himself more stringent and in some way more appropriate tolerance widths than were given by the visual and auditory cues. More specifically, the total score given at the end of a trial by the verbal cue may have helped *S* to define more effective standards of performance which varied with the stage of practice.

The presence of reference marks in the primary cue may have made possible this more effective use of the verbal cue at the same time that it minimized the importance of the visual and auditory cues. That the structure of the primary cue may be important in determining the role of augmenting cues is suggested by the fact that Goldstein and Rittenhouse (1954) using similar verbal and auditory augmenting cues in the Pedestal Sight Manipulation Test (PSMT) found that the verbal cue was inferior to the auditory cue during training. The primary ranging cue in the PSMT which is a pursuit tracking task is notably difficult to discriminate and lacks the reference marks provided in the present compensatory tracking task.

The present results may be compared to those obtained by Smode (1958) who also used a compensatory tracking task. He studied the effects of low- and high-level knowledge of results. His low-level condition corresponded approximately to our verbal condition. In his high-level condition he used discrete visual and auditory cues consisting of relay clicks and lights which counted each .5 sec. of cumulated time on target. Our auditory cue in a sense was more informative (since it was continuous and could be used to monitor performance at every instant) yet measured relative to the performance of the verbal group in each study, his auditory cue produced superior results. Nevertheless, these findings are not in conflict if we recognize that all of Smode's cues are essentially noncoincident cues. In comparison with the present study his results suggest that the efficacy of augmenting cues in the two types of task is mainly a result of their motivating properties as incentives and their help in defining standards of performance. For example, it is hypothesized that the clicks of Smode's study rewarded staying longer on target since they counted cumulative on-target time while the continuous cues of the present study may not have done this as well if at all.

These considerations may help to resolve the apparent conflict of results obtained in two other studies of augmenting feedback. Reynolds and Adams (1953) used auditory clicks at various intervals of delay to augment the primary cues of a pursuit rotor task and obtained positive results both for the effectiveness of delay as a variable and for the augmenting cue in general. Archer and Namikas (1953), interested primarily in the delay of feedback variable, tried to replicate this study except that they used a continuous tone instead of clicks as the augmenting cue. They obtained negative results and concluded that "delay of information feedback is not an effective variable." Bilodeau and Bilodeau (1961) in a recent review seem to agree with this conclusion. On the basis of the present analysis we would suggest that the failure to confirm the results of Reynolds and Adams was due to the change from a discrete to a continuous type of cue. It is felt that this change, contrary to the expectation of Archer and Namikas, may have reduced or eliminated the desirable incentive or reinforcing properties of the original cue and in this way reduced the effectiveness of feedback delay as a variable. Whether their augmenting cue was completely ineffective cannot be definitely determined since being interested only in the influence of varying delay in feedback they did not use a control group as did Reynolds and Adams.

On the basis of this discussion, then, we would conclude that augmenting cues are most effective in transfer when they function as incentives and in defining standards of performance; they are least effective in this respect when their major function is to provide information for guiding immediate action (Miller, 1953).

One other question of interest concerns the consistently poorer performance of the auditory cue compared with the visual cue. There is enough evidence here to warrant further research although these two cues differed in other respects so it is not possible at this point to attribute any difference specifically to a modality factor.

REFERENCES

ARCHER, E. J., KENT, G. W., & MOTE, F. A. Effect of long-term practice and time on target information feedback on a complex tracking task. *J. exp. Psychol.*, 1956, **51**, 103–112.

ARCHER, E. J., & NAMIKAS, G. A. Pursuit-rotor performance as a function of delay of information feedback. *J. exp. Psychol.*, 1958, **56**, 325–327.

BILODEAU, E. A. Some effects of various degrees of supplemental information given at two levels of practice upon the acquisition of a complex motor skill. *USAF HRRC res. Bull.*, 1952, No. 52-15.

BILODEAU, E. A., & BILODEAU, I., McD. Motor Skills Learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Rhinehart, 1960.

GOLDSTEIN, M., & RITTENHOUSE, C. H. Knowledge of results in the acquisition and transfer of a gunnery skill. *J. exp. Psychol.*, 1954, **48**, 187–196.

HOUSTON, R. C. The function of knowledge of results in learning a complex motor skill. Unpublished master's thesis, Northwestern University, 1947.

KARLIN, L. Psychological study of motor skills: Phase I. *USN Train. Dev. Cent. tech. Rep.*, 1961, No. 558-1.

KARLIN, L., & MORTIMER, R. G. Psychological study of motor skills: Phase II. *USN Train. Dev. Cent. tech. Rep.*, 1961, No. 558-2.

MILLER, R. B. Handbook on training and training equipment design. *USAF WADC tech. Rep.*, 1953, No. 53-136.

MORIN, R. E., & GAGNÉ, R. M. Pedestal Sight Manipulation Test performance as influenced by variation in type and amount of psychological feedback. *USAF HRRC res. Note*, 1951, No. MS 51-7.

REYNOLDS, B., & ADAMS, J. A. Motor performance as a function of click reinforcement. *J. exp. Psychol.*, 1953, **45**, 315–320.

SMODE, A. F. Learning and performance in a tracking task under two levels of achievement information feedback. *J. exp. Psychol.*, 1958, **56**, 297–303.

(Received February 1, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 80-85

STABILITY AND ADAPTATION OF SOME MEASURES OF ELECTRODERMAL ACTIVITY IN CHILDREN[1]

NORMAN L. CORAH AND JOHN A. STERN

Washington University School of Medicine

Measures of electrodermal activity in children, skin resistance levels and spontaneous fluctuations, made during rest and during tone stimulation were related to each other and to the conditions under which they were obtained. The results indicate that the resting measures are fairly reliable over time although they both showed the effects of adaptation, and they are influenced by the amount and type of prior activity. These measures showed significant but low relationships with each other. Orienting responses to tone stimulation tended to show the effects of adaptation from stimulation in a previous session.

This report is concerned with an analysis of two measures of electrodermal activity (EDA): resting skin resistance level and spontaneous fluctuations. The level of skin resistance presents few problems for conjecture in the present study since it has been shown to adapt in the laboratory (Duffy & Lacey, 1946), adaptation being manifested by increases in skin resistance.

The other measure has been referred to by various authors as "spontaneous activity" or "spontaneous fluctuations" (Burch & Greiner, 1960; Lacey & Lacey, 1958; Stern, Stewart, & Winokur, 1961). Spontaneous fluctuations (SF) are here defined as electrodermal fluctuations which are not related to specific experimentally induced stimulation.

The adaptability of the SF measure and its relationship to levels of skin resistance activity is subject to some controversy. Lacey and Lacey (1958) have expressed the viewpoint that spontaneous activity arises from physicochemical processes at the level of localized cell aggregates and is not merely responsivity to undetected external and internal stimulation. However, they do consider spontaneous activity as responsivity to "more generalized aspects of environmental demand" (p. 201). Their rationale for the conclusion that this activity is not related to specific responses centers around its reliability over time, the lack of significant adaptation obtained in their studies, and its low or insignificant relationship with levels of skin resistance.

The present authors would take the position that spontaneous fluctuations are responses to unknown stimuli, and as such, should show the usual effects of adaptation of autonomic measures. There are probably intraindividual differences in the number of SFs seen at "rest." However, under conditions of sleep or anesthesia we would suspect that SF would be reduced to zero in all *S*s. It would be expected that the amount of activity would decrease as adaptation progresses; i.e., skin resistance should increase and SF should decrease. We would also predict that SF should correlate significantly with the number of orienting responses obtained to specific stimuli if SF indicates responsivity of the electrodermal mechanism to stimulation. Stern et al. (1961) have demonstrated significant relationships between SF during tone stimulation and the number of orienting responses given to the tone. The present study differs from the earlier one in that it studies the same parameters in children as well as measuring the reliability of these measures.

---

[1] This work was supported, in part, by a research grant B 2449 from the National Institutes of Health, Public Health Service. The authors wish to express their thanks to Helen McMahon and Deborah Onken for their assistance in the collection and analysis of the data.

METHOD

Subjects.—Twenty-four children, 12 boys and 12 girls, between 7 and 8 yr. of age were obtained from a school near the laboratory. An additional 12 boys and 12 girls who were serving as controls in an anoxia study were also used. This second group was also 7 yr. of age. All children and their parents were paid for their participation.

Procedure.—Each child was brought into the laboratory and asked to rest in a hospital bed. Recording electrodes were attached to the first and third fingers of the right hand. These electrodes were made of zinc, approximately 15 mm. in diameter, backed with dental cement and were taped over corn pads filled with electrode jelly. This method was adapted from one recommended by Lykken (1959). The EDR recorder utilized a Wheatstone bridge with constant current supplied through the *S* (Grass 5 P1F preamplifier). The output of the bridge was fed into a dc chopper-type amplifier and continuously recorded on a Grass polygraph.

Each child was instructed to rest without moving. He was told that he would occasionally hear sounds coming from the speaker located approximately 10 in. over his head. First, a 10-min. recording of resting activity was made. Then, a series of 20 tones was presented over a period of approximately 15 min. The tones were produced by an Eico-Audio generator. The tones (500 cps, 27 db. above a background noise level of 40 db. spl) were all 6 sec. in duration. The intertrial interval was varied randomly between 20 and 60 sec. in 10-sec. steps. However, the interval between a given pair of trials was the same for all *S*s. The tone stimulation was followed by another resting period of 5 min. duration.

The first group of 24 *S*s repeated the procedure 24 hr. after their first experience with it. All of these *S*s appeared to be fresh and rested when they came to the laboratory. Half of this group, 6 boys and 6 girls, followed the above procedure on 2 successive days (Group IB). The other 12 *S*s (Group IA) received no tone stimulation on the first day and merely rested for 30 min. They received the regular procedure on their second day.

The second group of 24 *S*s (Group II) from the anoxia project were subjected to the regular procedure on only one occasion. In addition, the procedure was carried out after these *S*s had been involved in 2½ hr. of neurological, psychiatric, and psychological examinations on the same day.

The following measures were obtained:

1. Skin resistance. Levels of "apparent skin resistance" were taken at the end of each 2 min. of rest during the initial resting period and at the end of the third and fifth min. of the final resting period. However, some judgment was employed in taking these measures. If spontaneous fluctuations or movement-induced changes occurred at the end of the interval, the resistance was taken from the point just prior to this activity. The average resting resistance was the measure used. In addition, the resistance immediately prior to tone stimulation on Trials 5, 10, 15, and 20 was recorded and averaged for Groups IA and IB.

2. Spontaneous fluctuations. All fluctuations of 1,000 ohms or greater were tallied for each 2 min. of rest. The average number of SFs for the 2-min. periods was used for analysis. The number of SFs between trials of tone stimulation was tallied for Groups IA and IB. The average number of SFs per min. was the measure used.

3. Orienting response to tone. The measure consisted of the number of tone presentations to which an orienting response was given between 1 to 3 sec. (range) after initiation of the tone stimulus. The deflection had to be of an amplitude of 1,000 ohms or greater in order to be counted as a response. This measure was obtained for Groups IA and IB.

RESULTS

Stability of the measures.—These data were available from Groups IA and IB. All of the correlations presented are product-moment correlations. The correlation for average resting resistance between the first and second days was .86. The correlation for average number of SFs during rest on the 2 days was .61.

Within session data were available from the sessions on the second day. The correlations were .95 between average resting resistance and average intertrial resistance during tone stimulation, and .88 between average number of resting SFs and average number of intertrial SFs. All of these coefficients are significant beyond the .01 level.

Relationships among the measures.—The correlations between levels of skin resistance and SF are presented in Table 1. It may be noted that the only nonsignificant correlation obtained was in Group II.

The relationships between the responses to specific stimuli (orienting responses given to tone stimulation) and the other measures are given in Table 2. It should be noted that the combination of measures on which the correlations in the second column are based is a result of the fact that Group IA did not receive tone stimulation until the second day. The SF measures are consistently and significantly related to orienting responses to tone stimulation.

TABLE 1

Product-Moment Correlations ($N = 24$) between Levels of Skin Resistance and SF

Day 1

Resting Measures	Group I	Group II
	No. Resting SF	No. Resting SF
Resting resistance	−.51**	−.33

Day 2

Resting Measures	Group I		
	No. Resting SF	Intertrial Resistance	No. Intertrial SF
Resting resistance	−.46*	.95**	−.55**
No. resting SF		−.62**	.88**
Intertrial resistance			−.64**

$^{*}p < .05.$
$^{**}p < .01$

TABLE 2

Product-Moment Correlations ($N = 24$) between the Number of Orienting Responses to Tone Stimulation and the Other Measures of Activity

Resting Measures	No. Orienting Responses to Tones	
	Group I—Day 2	Group IA—Day 2 Group IB—Day 1
Intertrial resistance	−.35	−.53**
No. intertrial SF	.86**	.80**
Resting resistance	−.38	−.44*
No. resting SF	.74**	.58**

$^{*}p = .05.$
$^{**}p = .01.$

Adaptation of the autonomic measures.—The data for the resting measures are presented in Table 3. Both level of resistance and SF show significant adaptation effects from the first to the second days for Group I. In addition, the average number of SFs for the *final resting periods* (last 4 min. of each session) are presented for Group I. This difference is just short of significance at the .05 level.

The hypothesis that the effects of the preceding activities for the children in the anoxia project influences their autonomic measures was tested by comparing the resting measures for the first day from Group I with those from Group II (see Table 3). A t of 2.12 ($df = 46$, $p = .05$) was obtained for differences in average resting resistance of the two groups. A t of 1.57 was obtained for differences in average number of SFs which is not significant although the difference is in the predicted direction.

It was hypothesized that the effects of prior tone stimulation would be such as to increase the rate of adapta-

TABLE 3

Summary of Data Used to Test the Effects of Adaptation of the Resting Measures

Resting Measures	Group I Day 1			Group I Day 2			Group II Day 1	
	Mean	*SD*	t^a	Mean	*SD*	t^a	Mean	*SD*
Resting resistance (ohms)	241,851	81,261	—	275,330	97,446	3.29**	194,134	74,840
No. resting SF	10.87	6.39	—	7.32	4.96	2.87**	13.83	6.66
No. final resting SF	11.65	7.29	—	9.08	6.67	2.02		

Note.—$N = 24$.
[a] t test for correlated means.
** $p = .01$.

tion of the orienting responses to tone in the second set of trials. Hence, Group IB should show more rapid adaptation to tone stimulation on the second day than should Group IA. The results plotted by trials are illustrated in Fig. 1.

The hypothesis was tested by comparing the number of orienting responses to tone from the two groups. The data are presented in Table 4. Since the response tended to show relatively rapid adaptation under all conditions, the tests were also carried out for the number of responses in the first 10 trials only. The only result which is statistically significant is the change from Day 1 to Day 2 for Group IB although Fig. 1 shows a trend for Group IA to fall between the Day 1 and Day 2 results from Group IB. It must be noted that the size of each group is only 12 *S*s. Table 4 clearly indicates that the distributions are markedly skewed. However, nonparametric tests calculated on the data yielded essentially the same or lower levels of probability.

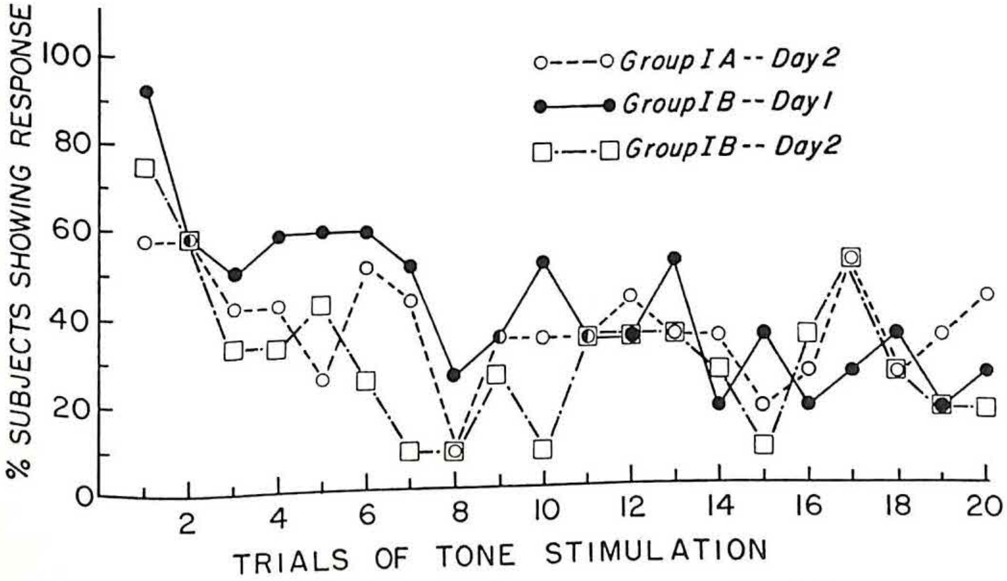


Fig. 1. Percentage of each group responding to tone stimulation.

TABLE 4

Tests of the Effects of Prior Experience with Tone Stimulation upon the Adaptation of Orienting Responses to Tone

No. of ORs to Tone	Day 1		Day 2		*t*
	Mean	*SD*	Mean	*SD*	
During 20 trials					
Group IB ($n = 12$)	8.25	4.20	5.92	4.59	1.82
Group IA ($n = 12$)	—	—	7.17	6.46	IB-Day 1 .58
During first 10 trials					
Group IB	5.33	2.53	3.17	2.69	2.48*
Group IA	—	—	3.83	3.38	IB-Day 1 1.23 IB-Day 2 .53

* $p = .05$.

Discussion

The data suggest that while resting measures of skin resistance and SF are fairly reliable over time, they show adaptation to particular situations and procedures. Hence, it can be expected that *S*s will maintain their relative ranks on these autonomic measures at least when the measures are taken under resting conditions.

The adaptation effect found for SF in the present and earlier studies (Stern et al., 1961) is not consistent with Lacey and Lacey's (1958) formulation concerning spontaneous activity as an intracellular effect which remains stable within the individual. Lacey and Lacey have demonstrated such stability when *S* is allowed to rest for a 20-min. period prior to recording such SFs. Under our experimental conditions, *S*s rested for approximately 10 min. prior to initiation of recording. These procedural differences might be invoked to account for the differences in the results obtained. However, the differences between the final resting SF, recorded at least 35 min. after *S* had begun to rest, from the first to the second day reaches the .06 level of significance. This result seems to be sufficient to warrant the retention of our conclusion.

Our results also indicate that the level of autonomic activity obtained in a given experimental session is influenced by factors external to the specific situation. The results from Group II suggest that activity occurring prior to the laboratory situation will influence the level of autonomic activity obtained in the laboratory.

Although the predicted trends concerned with differential adaptation of the orienting response to tone stimulation were present, they were not all significant. The small size of the groups has already been mentioned as one possible contributing factor. Another factor may be the intensity of the tone used. Comparison of the present results with others from our adult laboratories suggests that the children adapted more rapidly. The intensity of the tone above background noise is somewhat less than that used with the adults (c.f. Stern et al., 1961) and might thus be a contributing factor to the more rapid adaptation.

The writers tend to think of the measures used in this study as reflecting "activation" or the "drive state" of the individual (Malmo, 1959). That the frequency of SFs can be experimentally manipulated, and that they further are related to "alertness" as produced by pharmacological agents, has been demonstrated by Burch and Greiner (1960).

That the resting level of skin resistance is also affected by the level of "activation" is beyond doubt. As *S*s go from wakefulness to sleep, there is a marked increase in skin resistance; as *S*s are exposed to experimental stressors there are concomitant decreases in skin resistance. In the present study, these two measures of activation demonstrated statistically significant, though relatively low, relationships in two out of three comparisons made. In previous investigations using adult *S*s, we similarly found these two measures to be negatively though not significantly correlated with each other. The scatter plots in these earlier studies, involving larger samples than the present study, indicated that at low levels of skin resistance SF could fluctuate markedly between *S*s, some showing few, others many SFs. Why the correlations are higher for our population of children as compared to young adults is unclear.

In view of the relatively low correlations obtained between skin resistance and SF it would seem that these two are tapping different aspects of electrodermal activity. Further research is needed to clarify the similarities and differences in these measures.

REFERENCES

BURCH, N. R., & GREINER, T. H. A bioelectric scale of human alertness: Concurrent recording of the E.E.G. and G.S.R. *Psychiat. res. Rep.*, 1960, **12**, 183.

DUFFY, E., & LACEY, O. L. Adaptation in energy mobilization: Changes in general level of palmar skin conductance. *J. exp. Psychol.*, 1946, **36**, 437–452.

LACEY, J. I., & LACEY, B. C. The relationship of resting autonomic activity to motor impulsivity. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1958, **36**, 144–209.

LYKKEN, D. T. Properties of electrodes used in electrodermal measurement. *J. comp. physiol. Psychol.*, 1959, **52**, 629–634.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

STERN, J. A., STEWART, M. A., & WINOKUR, G. An investigation of some relationships between various measures of galvanic skin response. *J. psychosom. Res.*, 1961, **5**, 215–223.

(Received February 10, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 86–93

PERFORMANCE IN DIFFERENTIAL CONDITIONING AS A FUNCTION OF VARIATION IN MAGNITUDE OF REWARD

HENRY GOLDSTEIN AND KENNETH W. SPENCE

State University of Iowa

4 groups of rats designated Groups 10-0, 10-1, 10-5, and 5-1 were run in a differential conditioning experiment. The numerals refer to the number of food pellets (W_g) associated with the 2 discriminanda (white and black alleys). Ss ran 75 trials to each alley. Differentiation within an S occurred in groups rewarded in both alleys much as it did in Ss given no reward on 1 of the discriminanda. Rate of differentiation was positively related to differences in W_g. Comparison between groups showed that the function relating asymptotic response speeds for both large- and small-reward discriminanda was a negatively accelerated increasing one. Also, the finding that response speeds based on a given reward magnitude were the same whether it was the larger or the smaller of the 2 rewards implies that asymptotic performance is determined by the absolute magnitude of the reward and not by some other factor(s) based on contrast or relative effects, e.g., frustration.

Employing abstract theoretical constructs (intervening variables) initially inferred on the basis of the empirical laws of classical and instrumental conditioning, the learning theories of Hull (1952) and Spence (1956, 1960) are subsequently tested by deriving new implications concerning further phenomena to be expected either in these same situations or in more complex types of selective learning. While considerable progress has been made in this theoretical program, much remains to be accomplished. Especially deficient in the basic conditioning theory is lawful knowledge concerning the effects of differential reinforcement within a single *S*. Most of the studies of the variables known to determine response strength in conditioning have involved simple rather than differential conditioning. In simple conditioning experiments the effect upon behavior of varying a particular variable, e.g., reward magnitude, is investigated in terms of between-*S* differences. Such laws, however, do not provide us with knowledge as to the effects of within-*S* variation of the experimental variable upon performance, knowledge which is essential in connection with attempts to extrapolate the theory to more complex types of selective learning. Differential conditioning experiments are essential if we are to have available the laws describing such within-*S* effects.

Experimental studies of differential instrumental conditioning have, for the most part, been limited to the type in which response to one of the discriminanda (S+) only receives reinforcement, with response to the other (S−) receiving zero reinforcement. The findings of such studies (cf. Bower & Trapold, 1959; Friedes, 1957; Hunter, 1959; Raben, 1949; Spence, Goodrich, & Ross, 1959; Taylor, 1953) is that the speed of response curve to S+ increases in a negatively accelerated or S shaped fashion in a manner quite similar to that of an instrumental conditioning curve. The curve of response speed to S−, on the other hand, at first rises along with the S+ curve, but then after a number of trials diverges from it and either ceases to rise or decreases to a lower level.

Investigation of the magnitude of reward variable (W_g) in such differential instrumental conditioning has also been limited primarily to employing different reward amounts with the positive discriminandum, zero reward being given on the negative. Two such studies (Bower & Trapold, 1959; Hunter, 1959) have both reported that, as in the case of simple instrumental conditioning, speed of response to the rewarded S+ was a positive function of W_g. In the case of response to the negative, nonrewarded alley (S−), there was no difference in the final speeds (at the end of training) for the *S*s run with different reward magnitudes on S+. It is thus apparent that the degree of differentiation was also related to W_g.

The present study was interested in extending the study of differential instrumental conditioning to the situations in which responses to both discriminanda are reinforced, but with different reward magnitudes. Four groups, designated 10-0, 10-1, 10-5, and 5-1 were run in the experiment. The numeral before the hyphen indicates, for each group, the number of pellets received following response to the stimulus (arbitrarily designated S+) which led to the larger reward magnitude. The numerals after the hyphen indicate the number of pellets received upon responding to the stimulus (S−) which led to the smaller reward magnitude. Such an experimental design provides information on a number of different points:

1. Whether or not rats are able to learn to respond differentially to discriminanda, responses to which receive differential magnitudes of reward. Related to this question is that of whether the degree of differentiation is related to the size of the differences between the reward values.

2. How performance to the larger rewarded discriminandum (S+) varies (*a*) as a function of the magnitude of reward on S+, with reward to S− constant; (*b*) as a function of reward magnitude on S−, with reward to S+ equated.

3. How performance to the smaller rewarded discriminandum (S−) varies as (*a*) a function of the magnitude of reward on S− with reward to S+ equated; (*b*) as a function of reward magnitude on S+, with reward to S− equalized.

4. Whether or not performance associated with a given reward magnitude is dependent on the value of the reward on the comparison discriminandum. For example, how will the performances of *S*s in Groups 10-5 and 5-1 to the discriminandum rewarded with 5 food pellets compare? As may be seen in the case of one group, 5 pellets is the larger reward while in the other it is the smaller.

METHOD

Subjects.—The *S*s were 28 male and 28 female, experimentally naive hooded rats from the Iowa Psychology Department colony. They ranged in age from 150–180 days at the beginning of pretraining.

Apparatus.—The apparatus consisted of two parallel straight wooden alleys, each 60 in. in length, 4 in. in height, and 3 in. wide. The first 10 in. of each alley served as a start box. Each was separated from the remainder of the alley by a single, downward-opening aluminum door which served as a start door for both alleys. Two appropriately painted, metal retrace doors partitioned off the last 15 in. of each alley to form goal boxes. To the rear wall of each goal box was fastened an aluminum food cup painted the appropriate color.

The apparatus was placed in a dimly lighted room and was completely covered and surrounded by black cloth in order to provide a homogeneous surrounding environment.

Each start box was painted a flat medium gray as was the side of the start door facing the start boxes. One runway and its goal box were painted a flat white and the other runway and its goal box were painted a flat black. The alleys were illuminated by two 40-w. fluorescent tubes, 4 ft. long, one tube being centered 40 in. above the longitudinal axis of each alley floor. Under the conditions employed, the alleys had the following reflected brightnesses: white, .08 mL.; black, .008 mL.; gray, .033 mL.

The timing circuits consisted of Hunter Klockkounters, Hunter silent contact relays, and photocells set in the side walls of the alley. The light sources for the photocell beams were infrared filtered. The circuits measured response times in three portions of the alley. The first interval, initiated with the opening of the start door was terminated when *S* interrupted the first light beam located 6 in. from the start door. The second interval began with the end of the first one and ended at a distance 6 in. beyond the first. The third interval began with the end of the second and ended with the interruption of the final beam 36 in. beyond the second one. The latter corresponded to a position just above the food cups. All time data were converted to speed measures in terms of ft/sec. The three measures will hereafter be referred to as starting speed, running speed, and goal speed, respectively.

Conditioning procedure.—Two weeks before the beginning of experimental training, *S*s were removed from the colony room and placed in home cages in a room that was kept in total darkness except when *E* was present for purposes of feeding the *S*s, removing them from their home cages to the transport cages, or cleaning the home cage racks. Throughout this period *S*s were allowed access for 1 hr. per day to a wet-mash mixture consisting of approximately 50% dry mash and 50% tap water by weight. Drinking water was always available. This regimen was continued until completion of the experiment.

During this 2-wk. period *S*s were also accustomed to being handled and placed in the carrying cages that were to be used to transport them from the home cage room to the experimental room. The *S*s were given no pretraining experience of any kind in the experimental apparatus. The temperature of both the home cage and experimental room was maintained at 74–76° F.

Procedure.—For all *S*s the experimental design was such that each received one magnitude of reward (a certain number of food pellets[1]) after traversing an alley of one brightness (e.g., white) and a different amount of reward after traversing the alley of the opposite brightness (e.g., black). As described in the introduction, four groups of *S*s, designated Group 10-0, Group 10-1, Group 10-5, and Group 5-1 were run in the experiment. Each group was composed of half females and half males. Within each group the reward magnitude—brightness conditions were counterbalanced with respect to sex.

The experiment was conducted in two independent replications, with one group (5-1) not represented in the second replication. There were 8 *S*s in each group of a replication.

For the first 4 days of each replication only 3 trials per day were administered. Thereafter each *S* was given 6 trials per day, 3 to S+ and 3 to S−. The order in which positive and negative trials were administered within each daily block of 6 trials was as follows: The first trial was always positive and the second was always negative. The last 4 trials of each day, which provided the data reported in this paper, were ordered by repeatedly cycling through a fixed sequence of 16 trials constructed so as to insure equal numbers of positive and negative trials for each day and a limit of 2 similar trials in succession both within a day and from one day to the next. The sequence used for these last 4 trials of each day was − + − +, + − + −, − + + −, + − − +. This procedure of holding the order of the first 2 trials of each day constant and analyzing only the measures obtained during the last 4 daily trials was adopted with a view to eliminating the "warm-up effect" per se and the day-to-day variability associated with the use of varying sequences of reinforcements in the warm-up trials.

The *S*s in each of the experimental replications were divided into four squads, each squad consisting of one male and one female from each of the experimental groups. The *S*s within a given squad completed all of their daily trials before the next squad was begun. The order in which the four squads were run from day to day was kept constant, and within any one squad the reward magnitude-brightness conditions were the same for all members, i.e., for all members of Squads 1 and 4, white was S+ and black was S−, whereas for all members of Squads 2 and 3, black was S+ and white was S−. Furthermore, the various experimental groups were represented in the same order from squad to squad and from day to day, i.e., first the males were run in the order 10-0, 10-1, 10-5, 5-1 and then the females were run in the same order.

---

[1] The reward consisted of small round pellets weighing approximately 45 mg. each, obtained commercially from the P. J. Noyes Company, Lancaster, New Hampshire.

The procedure for any one squad involved having all *S*s complete a given trial (e.g,. the first) before the next trial was begun. This procedure resulted in a 6–8 min. intertrial interval during the first replication (8 *S*s per squad) and a 5–7 min. intertrial interval during the second replication (6 *S*s per squad). Approximately 15 min. after the members of the last squad were returned to their home cages all *S*s were allowed 1 hr. of access to a wet mash mixture.

Each individual trial began with the placement of *S* in the start box leading to the appropriate alley. As soon as *S* remained oriented toward the start door for approximately 2 sec., the door was opened. After *S* had broken the final photobeam, the retrace door was closed. The *S* was removed from the goal box *immediately* after consuming the reward pellets. If no reward pellets were present in the food cup (as was the case on half of the trials for the *S*s of the 10-0 group) *S* was removed *immediately* after the breaking of the final photobeam. It was decided, beforehand, to adopt a time-limit criterion on a trial of 120 sec., i.e., if after *S* left the starting chamber, it remained in any section of the alley for 120 sec., it was removed from that section and placed at the food cup if pellets were present. If no pellets were in the food cup *S* was removed from the apparatus. On such trials a latency of 120 sec. was assigned for those measures corresponding to portions of the alley which had not been fully traversed during the trial. This procedure was followed throughout a total of 150 trials (75 to each discriminandum) in each replication.

RESULTS AND DISCUSSION

Course of differentiation.—A picture of the manner in which the performance differential based on different reinforcement magnitudes develops within *S*s is provided in Fig. 1. This graph presents the running speed measures to the two brightness discrimininanda as a function of successive blocks of training trials for the three groups in which the larger reward was 10 pellets. As may be seen, differentiation in the case of Groups 10-1 and 10-5, in which reward was present in both alleys, develops in much the same manner as that for Group 10-0 in that, initially, the S+ and S− curves rise together for a number of trials and then the S− curves break away from the S+ curve, remaining approximately at the break-off level thereafter. This, it will be recalled, is essentially the same finding as was reported in previous curves of such standard differential instrumental conditioning cited earlier.

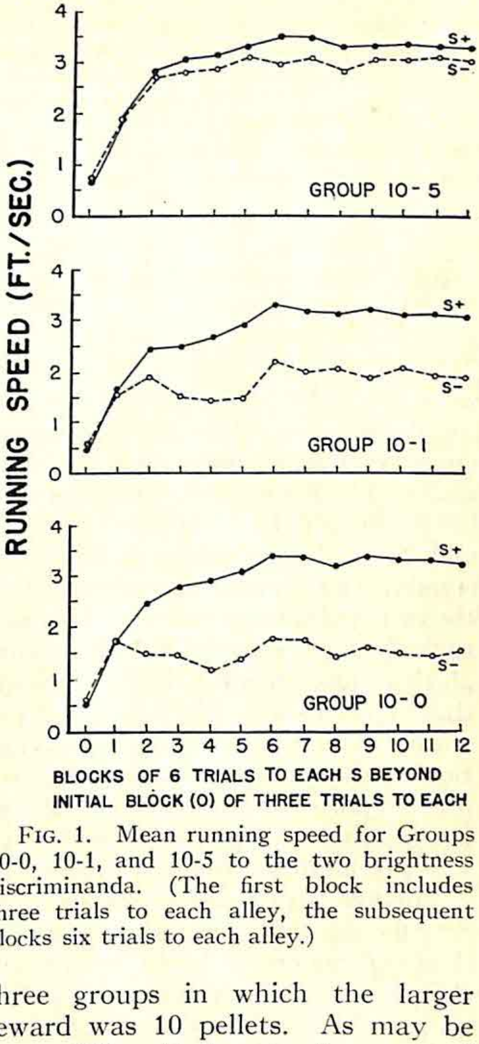


FIG. 1. Mean running speed for Groups 10-0, 10-1, and 10-5 to the two brightness discriminanda. (The first block includes three trials to each alley, the subsequent blocks six trials to each alley.)

The results for the starting speed measure were practically identical

TABLE 1

COMPARISON OF MEAN ASYMPTOTIC RESPONSE SPEEDS (FT/SEC) TO DIFFERENTIALLY REINFORCED DISCRIMINANDA FOR GROUPS WITH DIFFERENT COMBINATIONS OF REWARD MAGNITUDES

Response Measure	Groups Designated by No. of Pellets to Each Discriminandum							
	10 vs. 0		10 vs. 1		10 vs. 5		5 vs. 1	
Starting	1.44	.83	1.39	.90	1.50	1.35	1.45	1.18
Running	3.26	1.54	3.10	1.98	3.28	3.07	3.30	2.57
Goal	2.51	.97	2.34	1.89	2.70	2.61	2.60	2.12

with those obtained with the running speed measure, while the goal-speed measure differed only slightly. The major difference in the case of the latter occurred in Group 10-0, in which the curve for S− exhibited a steady, continuous decrement after the peak point was reached. This was the only S− curve of the 12 involving all three measures for the four groups that showed such a decrease. Earlier studies involving differential conditioning with no reinforcement on S− have reported both results. As yet the conditions responsible for the two kinds of findings are not known.

Further examination of Fig. 1 reveals that differentiation tends to develop sooner the larger the reward differences. Thus the curves for S+ and S− broke away after 18 trials in the case of Groups 10-0 and 10-1, but only after 30 trials in the case of Group 10-5. Not represented in the graph are the curves for Group 5-1, which required 42 trials before separation occurred.

Performance at asymptote.—As Fig. 1 shows, all of the curves for the larger reward alley appear to have reached asymptotes by the end of the acquisition period. More complete data on the performance levels at this stage for all four groups are presented in Table 1. This table gives the means of the response speeds for all three measures to the larger rewarded (S+) and smaller rewarded (S−) discriminandum for all four groups over the last 12 tabulated training trials (6 to S+ and 6 to S−). As may be seen, in every instance performance to S+ was higher than to S−. However, only in the case of Groups 10-0 and 10-1 were the differences in speed of response to the two discriminanda for all three measures significant $(p < .01)$.[2] For the 10-5 group the difference between S+ and S− asymptotic performance reached the .05 level in the case of one measure, starting speed. A separate analysis for the 5-1 group[3] revealed that none of the measures provided evidence for differentiation even at the .05 level of significance. It will be recalled that only one replication (8 *S*s) was run with this group. Possibly significant differentiation would be obtained with these relative reward magnitudes with more *S*s.

---

[2] These comparisons were obtained as simple effects from separate mixed analyses of variance of asymptotic performance (Trials 133–150) on each speed measure. Each such analysis included groups (10-0, 10-1, and 10-5), sex, and the counterbalanced factors of brightness and discriminanda as main factors. The simple discriminanda effects within each group were tested against the within-*S*s error term from the corresponding overall analysis.

[3] The analyses for Group 5-1, one for each speed measure, were identical with those described in Footnote 2, except that they were restricted to the single group.

Turning from these within-*S* differences to differences in performance between groups, attention is first directed to the fact that asymptotic performance to an S+ reward magnitude of 10 did not differ with variations of the smaller reward magnitude. Thus, as Table 1 reveals, Groups 10-0, 10-1, and 10-5 all responded at approximately the same level to the 10-pellet alley in the case of all three response measures. The small differences obtained are well within the range of variation to be expected by sampling. It would thus appear that performance at the asymptote to this large reward discriminandum is not affected by the magnitude of the reinforcer given on the alternative small reward discriminandum and hence the speed of response to the latter.

Looking next at the asymptotic speeds to the small reward discriminanda, it is apparent from the data for the different measures given in Table 1 that there is a consistent rise in the response speeds to reward magnitudes of 0, 1, and 5 pellets in the case of the three groups in which the larger reward magnitude was 10 pellets. The form of this function for the running speed measure is revealed by the dotted curve joining the open circles of Fig. 2.

Also presented in Fig. 2 are the results for the running speed measure to the large reward discriminanda. This (solid) curve joining the solid dots was determined not only from the data of the present experiment, but also from a study conducted concurrently with the present one that used the same apparatus and the

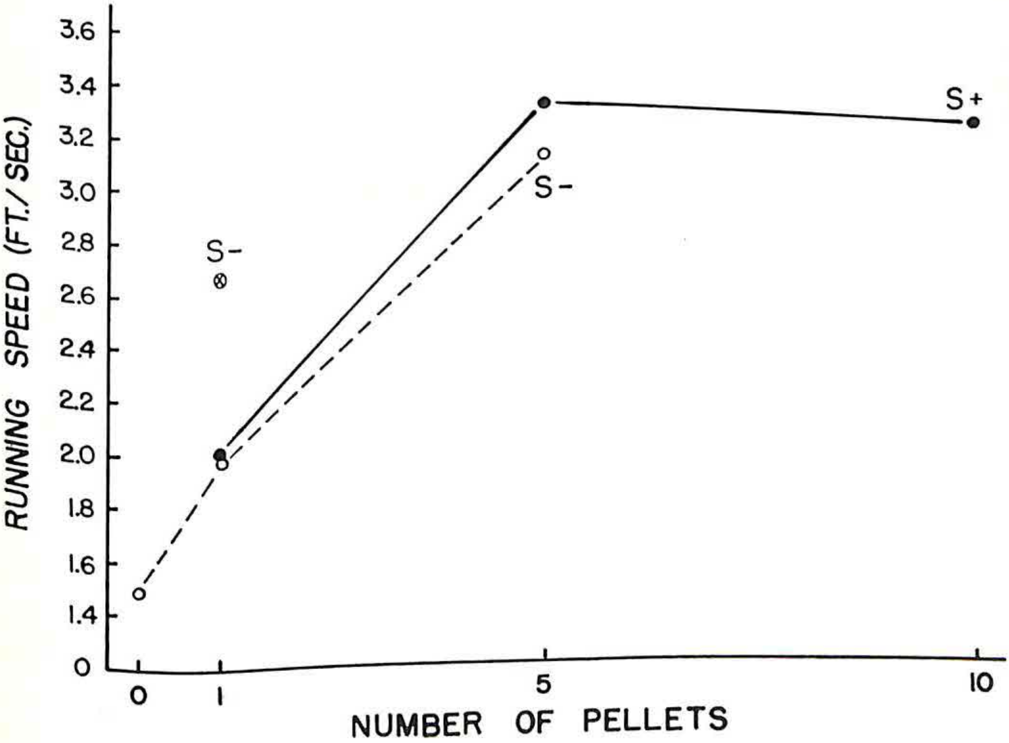


FIG. 2. Relation between mean running speed and number of reward pellets for both S+ and S−.

same procedures with respect to pretraining, motivation, and conduct of an experimental trial (Trapold, 1961). Thus the point on this curve for the reward value of one pellet was obtained from Trapold's Group 1-0.[4] The speed measure at 5 pellets was obtained from our 5-1 group, while the measure at 10 pellets represents the mean speed to the large reward discriminandum for our Groups, 10-0, 10-1, and 10-5. Also shown in the graph by means of the circled cross is the mean asymptotic running speed for Group 5-1 to the discriminandum that received 1 pellet.

Like the broken curve, which it follows closely at its common abscissa values, the solid curve suggests a negatively accelerated relation between asymptotic response speed and reward magnitude generally similar to those which have been obtained in simple instrumental conditioning studies in which the complication of differentiation is not present (cf., Crespi, 1944). The closeness of the curves at the common reward values means, of course, that response speed based on a given reward magnitude, is approximately the same to a "larger" rewarded discriminandum as it is to a "smaller" rewarded one. This implies that asymptotic performance to a particular discriminandum is determined by the absolute magnitude of the reward received and is not a function of its relative magnitude, i.e., the relation of this magnitude to the magnitude of reward given on the other comparison discriminandum.

This finding has very important theoretical implications for it suggests that the differential strengths of the response tendencies (E) in such differential instrumental conditioning is, at least *at the asymptote*, entirely a function of the incentive motivation variable (K) and is not due to some other factor such as frustrative inhibition (I) in the case of the smaller reward (cf., Spence, 1960). That is, a theoretical possibility that could be entertained is that a reward magnitude of 1 pellet, when given in connection with a 10-pellet reward, would result in frustration and hence some kind of an inhibitory (decremental) factor which, in turn, would have reduced the effective strength of the response to this discriminandum. The difference in the strengths of the responses to the two discriminanda would, in this case, have been due to differences in I as well as differences in K. Moreover, it would be predicted by such a theory that the asymptotic performance level to a discriminandum reinforced with a particular reward magnitude, e.g., 2 pellets, would be lower when this reward size was paired with a larger reward than with a smaller. The evidence of the present experiment, as we have seen, is against such a theory.

Attention should be directed to the procedure employed in the present experiment of removing an S from the goal box immediately after it picked up the pellets from the food container. It is possible that this procedure did not provide an opportunity for frustration to occur in the goal box and hence would account for the absence of such frustration effects with the smaller of the two rewards.

Further support for the interpretation that *asymptotic* performance level to two differentially reinforced discriminanda is, under the conditions of the present experiment, solely a function of incentive motivation, K, as determined by the absolute magnitude of each reward, is found in certain further data provided by Trapold's (1961) study. In addition to training groups under differential reinforcement, Trapold had one group which received 10 pellets on both brightness discriminanda and another group which received 1 pellet on each of the discriminanda. According to the interpretation that K is the primary determiner of the asymptote, these groups should have performed at the same level as the groups in the present experiment did to the discriminandum on which they received these particular reward magnitudes. Confirmation of this implication is seen in the fact that Trapold's group ($N = 16$) that had 1 pellet on both discriminanda achieved an asymptotic running speed of 1.95 ft/sec, which is almost exactly the same as the 1.98 ft/sec value attained by Group 10-1 of the present experiment on the discriminandum receiving 1 pellet. Again, Trapold's Group 1-0 ($N = 16$) reached an asymptotic level of 2.07 ft/sec on the larger rewarded alley which differs negligibly from the other two speeds for this reward amount (i.e., 1 pellet).

---

[4] The mean speed to the zero rewarded discriminandum of Trapold's 1-0 group, not represented on the graph, was exactly the same as that of Group 10-0 for the zero rewarded discriminandum.

REFERENCES

BOWER, G. H., & TRAPOLD, M. A. Reward magnitude and learning in a single-presentation discrimination. *J. comp. physiol. Psychol.*, 1959, **52**, 727–729.

CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

FRIEDES, D. Goalbox cues and pattern of reinforcement. *J. exp. Psychol.*, 1957, **53**, 361–371.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

HUNTER, J. J. Performance in differential conditioning as a function of magnitude of reward. Unpublished master's thesis, State University of Iowa, 1959.

RABEN, M. W. The white rats discrimination of differences in intensity as measured by a running response. *J. comp. physiol. Psychol.*, 1949, **42**, 254–272.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. The roles of reinforcement and nonreinforcement in simple learning. In, *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

SPENCE, K. W., GOODRICH, K. P., & ROSS, L. E. Performance is differential conditioning and discrimination learning as a function of hunger and relative response frequency. *J. exp. Psychol.*, 1959, **58**, 8–16.

TAYLOR, J. E. The effects of positive-negative cue differences on response tending. Unpublished doctoral dissertation, State University of Iowa, 1953.

TRAPOLD, M. A. Mixed- and separate-phase differential conditioning as a function of magnitude of reward. Unpublished master's thesis, State University of Iowa, 1961.

(Received February 10, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 94–102

STIMULUS GENERALIZATION IN THE LEARNING OF CLASSIFICATIONS

ROGER N. SHEPARD[1] AND JIH-JIE CHANG

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

6 different classifications of the same set of 8 Munsell colors into 2 equal subsets were learned by a paired-associate method in which one verbal response was associated with each of the 4 colors in one subset while a 2nd verbal response was associated with each of the 4 colors in the other subset. A condition of identification learning was also included in which a different response was associated with each of the 8 colors. In accordance with a proposed principle of stimulus generalization for classification learning, 78% of the variance of the difficulties of the 6 classifications was found to be accounted for by the pair-wise confusions among the same stimuli as determined from identification learning.

One purpose of the present study is to clarify the relation between two kinds of paired-associate arrangements. One, called the *identification learning* arrangement, conforms to the conventional paired-associate paradigm. Specifically, in this study, a one-to-one association is built up (by the usual method of anticipation) between eight individually presented stimuli and eight arbitrarily assigned verbal responses. The other, called the *classification learning* arrangement, departs from this paradigm in just this one respect: the number of responses is reduced from eight to two and one of these two responses is associated with four of the eight stimuli while the other response is associated with the remaining four stimuli.

Since the responses are essentially arbitrary verbal labels, the difficulty of learning any particular classification should not depend upon which of the two possible assignments is enforced between the two responses and the two subgroups of stimuli. Consequently the number of different classifications that must be distinguished here is just the number of distinct partitions of the set of eight stimuli into two equal subsets; viz., 35. These 35 classifications will themselves generally vary in difficulty, however. For example, to learn one name for each of four horses and another name for each of four dogs would surely be easier than to learn the one name for two of the horses and two of the dogs and the other name for the remaining two horses and dogs.

One factor that has been proposed to account for such variations in difficulty is that of stimulus generalization (Baum, 1954; Buss, 1950; French, 1953; Gibson, 1940; Newman, 1956; Oseas & Underwood, 1952). The above example might be understood in terms of this factor by noting that a response associated with any one of the eight animals should generalize most extensively to those stimuli that are psychologically the most similar to it; namely, the three other animals of the same species. Indeed, the experiment by Baum (1954) has already indicated that some of Heidbreder's (1947) results on the learning of concepts might be explained in this way. Still, the extent to which variations in difficulty between different classifications can be given a detailed, quantitative account in terms of the principle of stimulus generalization alone has not been precisely determined owing, perhaps, to the absence of a sufficiently rigorous and operational formulation of that principle.

---

[1] I am indebted to the late Carl I. Hovland for a number of helpful suggestions concerning an earlier version of this paper. Indeed I wish to use this opportunity to acknowledge more broadly that, from the time when I first became his student until his untimely death, Carl Hovland benefited me more than I can express, by his encouragement, by his advice, and—above all—by his example.

As first author of the present report, I am responsible for the conception and design of the experiment as well as for the report itself. The second author carried out the experiment proper and much of the tabulation and analysis of data. The stimulus materials were prepared by John Gibbon.

It was for just this reason that an explicit and quantitative interpretation of the principle of stimulus generalization, which had previously been proposed for identification learning (Shepard, 1957), was recently extended to classification learning as well (Shepard, Hovland, & Jenkins, 1961, pp. 25–29). According to this interpretation, if the responses are sufficiently distinctive that they are not themselves confused, then the overt errors observed during identification or classification learning alike should be attributable solely to the pair-wise confusions between the stimuli. Under this condition, then, the frequency with which the response assigned to one stimulus is incorrectly made to some other stimulus should depend only upon the psychological similarity between those two stimuli, and not at all upon any other aspect of the assignment (such as the total number of alternative responses in particular).

Unlike the earlier formulations of the principle of stimulus generalization, this formulation provides for a complete, quantitative prediction of the total number of errors to be expected during the learning of any one of the 35 different classifications. These predictions can all be based upon the results of a single experiment on identification learning using the same eight stimuli. Indeed, from the matrix giving the frequency with which each of the eight responses was made to each of the eight stimuli during identification learning, a prediction can immediately be derived not only for the total number of errors for any given classification but also for the number of incorrect responses for each stimulus in that classification taken separately. The predicted number of errors for Stimulus S_i is obtained from the matrix simply by summing those four entries in Row i that represent confusions between S_i and the four stimuli assigned to the other response in the given classification. (Clearly, confusions between S_i and stimuli assigned to the *same* response as S_i will not contribute to the overt errors.)

This formulation has already been subjected to a partial test in the recent investigation of classification learning by Shepard, Hovland, and Jenkins (1961). The stimuli in that investigation were psychologically highly analyzable in the sense that they tended to be immediately perceived or described in terms of their values on a small number of perceptually isolated and salient dimensions (such as size, color, and shape). The results conclusively showed that, for such analyzable stimuli, the actual difficulties of the different classifications departed in a very striking manner from the difficulties predicted from the proposed formulation of the principle of stimulus generalization. The failure of this prediction was attributed, in that study, to the intervention of an additional process of selective attention or abstraction which, it was argued, would be particularly likely with such highly analyzable stimuli.

The primary purpose of the present study, then, is to test this explanation of the earlier predictive failure by attempting the prediction anew with much more internally homogeneous or unanalyzable stimuli; namely, eight Munsell colors. Since a patch of uniform color tends to be reacted to as a whole (rather than analyzed into component parts), the proposed formulation of the principle of stimulus generalization alone might in this case be expected to account for the observed differences in difficulty among the various classifications.

METHOD

Subjects.—The *S*s were 38 female high school graduates employed in a clerical capacity by the Bell Telephone Laboratories.

Stimuli.—The stimuli were prepared from Munsell sheets of the same nine shades of red previously used in a study of stimulus generalization in identification learning (Shepard, 1958). Since the present experiment calls for eight rather than nine stimuli, one of these colors (S_4) was arbitrarily eliminated for this experiment. However, in order to facilitate direct comparison with the previous experiment, the eight remaining stimuli will still be denoted by the symbols used in the earlier experiment; viz., S_1 through S_9, omitting S_4. According to their Munsell designations these eight colors are all of the same red hue (viz., 5R) but vary in brightness and saturation with the following Munsell value|chroma specifications (arranged in order from S_1 through S_9): 7|4, 6|6, 6|10, 5|8, 5|12, 4|6, 4|10, and 3|4 (Munsell Color Company, 1945). Each stimulus consisted of a homogeneously colored disk (3/4 in. in diameter) centered on a white card. A description of the method of preparation and presentation of the stimuli, as well as of the precautions taken to insure that responses would be learned to the colors per se (rather than to any incidental cues) can be found in the earlier report. We only remark, here, that 16 cards were prepared from each of the eight color sheets and that, as a consequence, no *S* saw any one card more than twice during a learning session.

In order to illustrate the various classifications, it is convenient to represent the eight stimuli as eight points in a color space in such a way that the distance between two points is a decreasing function of the perceived similarity between the two corresponding colors. Such a representation, furnished by the multidimensional scaling solution obtained in the earlier study (Shepard, 1958), is shown in Fig. 1. About 99% of the variance of the frequencies of confusions between stimuli during identification learning in that study is accounted for by the distances between the points in this plot.

Classifications.—Six of the 35 distinguishable classifications were selected for systematic study. These were chosen both to cover the widest possible range of predicted difficulty and to be mutually orthogonal. (We call two classifications "orthogonal" if the four stimuli that constitute either subclass in one of the classifications split equally into the two subclasses in the other classification.) The multidimensional scaling solution of Fig. 1 is used in Fig. 2 to illustrate each of the 6 chosen classifications by coloring the four points corresponding to the stimuli associated with one response black, and by leaving the remaining four points white. According to

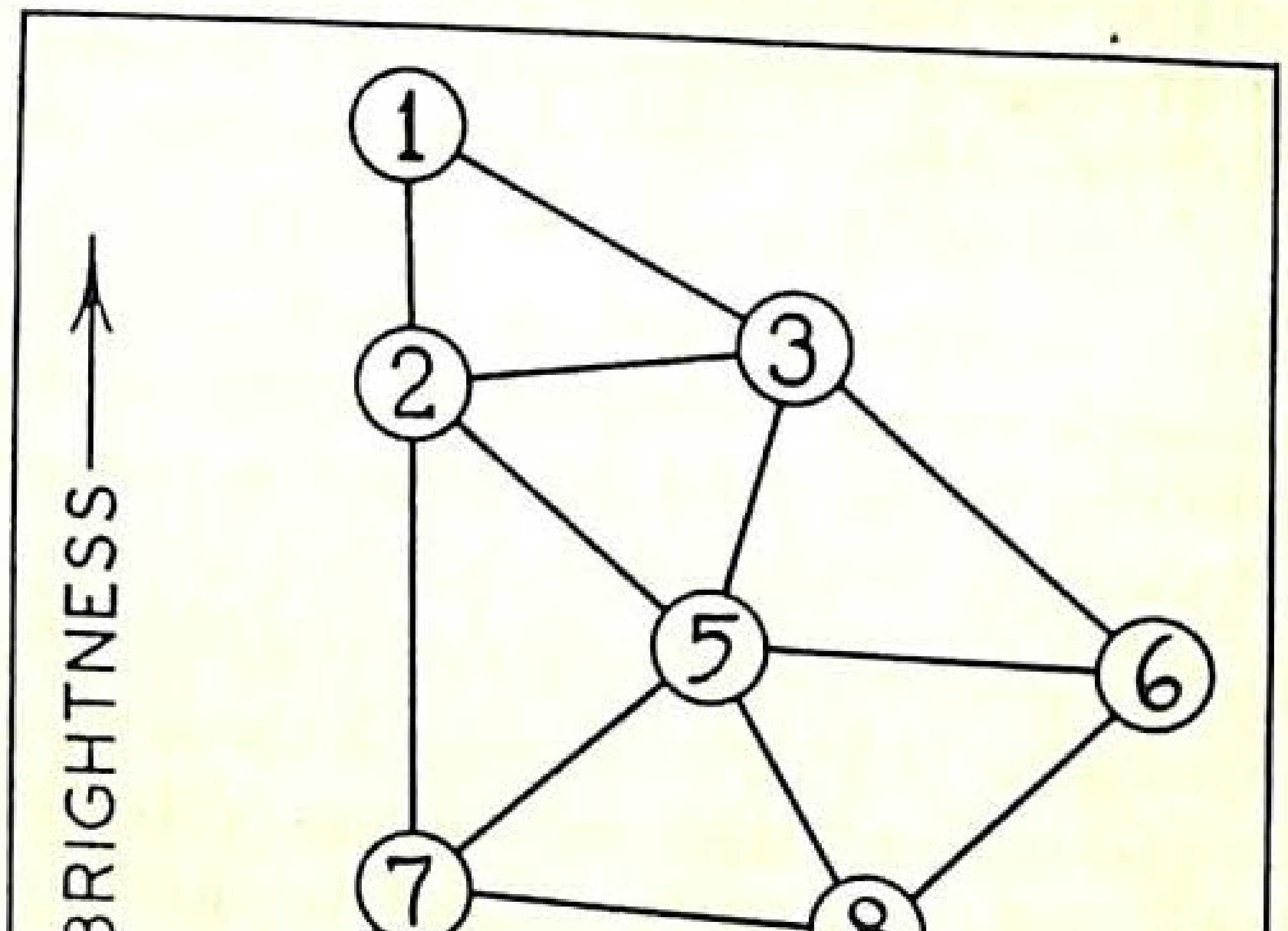


FIG. 1. A two-dimensional plot of the eight stimuli. (The smaller the separation between two points in this plot, the greater is the expected frequency of confusion between their corresponding colors.)

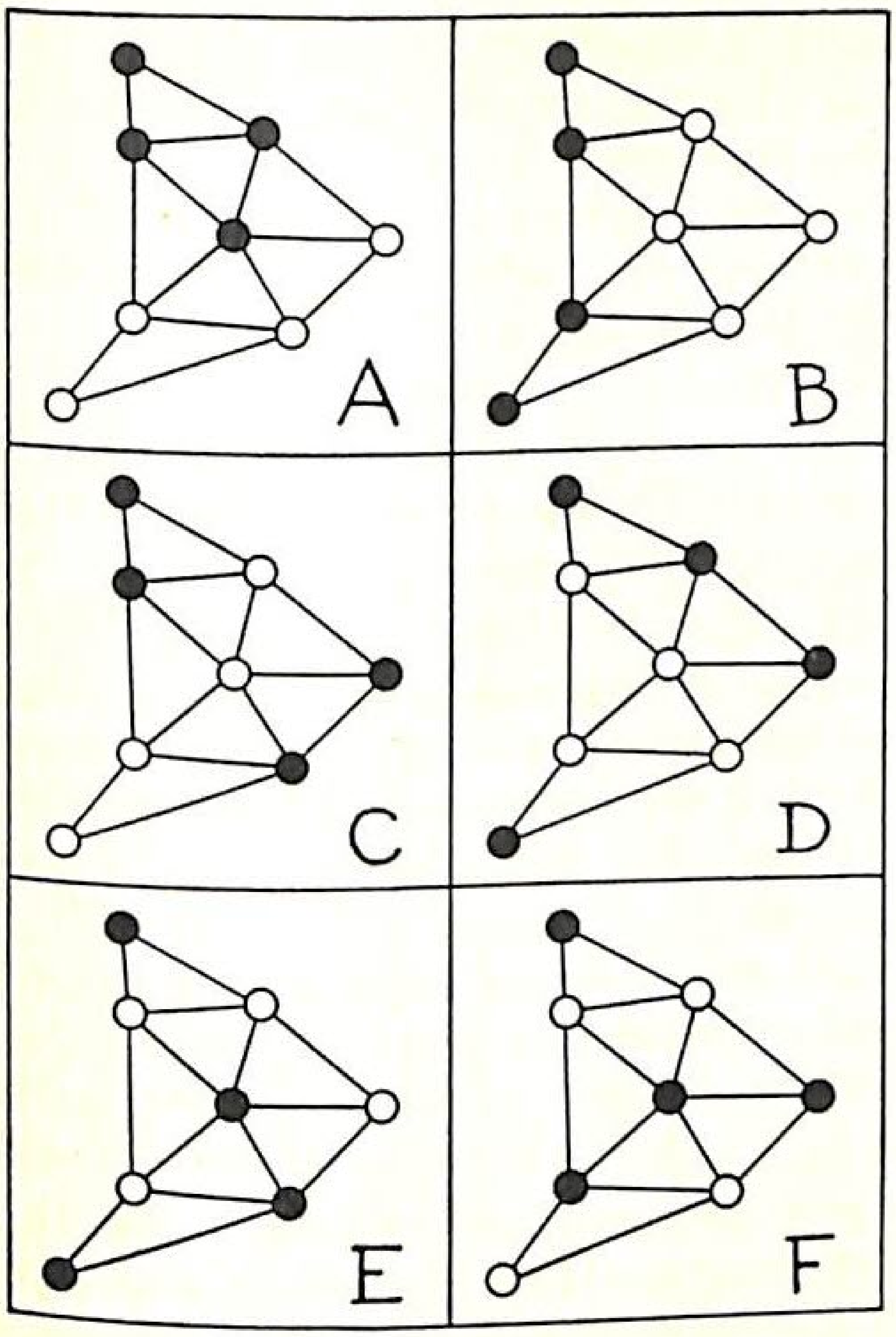


FIG. 2. Six orthogonal classifications of the eight colors. (The number-names for the specific colors included in either subclass of any of the six classifications can be found by comparing the appropriate plot with Fig. 1.)

our proposed formulation of the principle of stimulus generalization, classifications (like A) in which the black points are clustered together in one region and the white points are clustered together in another should be easier to learn than classifications (like F) in which the black and white points are intermingled.

Learning task.—The task was essentially the same as that employed in the earlier experiment on identification learning, but the number of responses was reduced to two in order to convert the identification task into a classification task. Each *S* was run individually, and stimuli were presented one at a time in random order. During any given learning session *S* was required to produce one of two specified verbal responses (e.g., "plus" or "minus") during the presentation of each stimulus (beginning with the first). The association between each stimulus and its assigned response was built up by the usual method of anticipation (or response correction) but the task was self-paced; i.e., *E* waited until *S* had responded to a given stimulus before stating the correct response, and then presented the next stimulus. Learning was continued for 200 stimulus presentations or until *S* had made 32 consecutive correct responses—whichever occurred first. The deck of 8 × 16 or 128 stimulus cards was reshuffled after each individual session.

Experimental design.—Owing to limitations on the number of available *S*s, it was found necessary to have each *S* learn more than one classification (hence the requirement that the classifications be orthogonal). Initially, then, each of 30 of the *S*s was given two different classifications—one on each of 2 consecutive days. In order to counterbalance possible order effects, each of the 30 possible sequences of two different classifications was used just once. Since there was a rather large variation in the performance of different *S*s on the same classifications, an attempt was made to secure more reliable data by replicating the entire procedure on the same 30 *S*s about 2 wk. later. The order of administration of problems was counterbalanced as before but with the additional constraint that none of the six classifications was given to any *S* twice. Thus each *S* learned exactly four different classifications. The two response alternatives were A and B, for the first classification; + and −, for the second; X and Y, for the third; and 1 and 2, for the fourth.

The test of the proposed explicit interpretation of the principle of stimulus generalization was to consist in a comparison of the number of errors made on each of the six classifications with the number predicted on the basis of the results obtained from the 36 *S*s of the earlier experiment on identification learning. However it seemed probable that the predictions would need to be adjusted in order to accommodate any overall difference in motivation or learning skill between the old and new populations of *S*s. (The earlier *S*s were primarily graduate students at Yale University, whereas almost all of the present *S*s had never attended college.) In order to estimate the size of the adjustment needed, a small sample of the present population of *S*s (consisting of the remaining 8) was given an identification learning problem with the same eight stimuli. This identification problem was administered in exactly the same way as the classification problems except that there were eight responses instead of two. These responses (viz., D, H, K, M, O, R, S, and W) were assigned to the stimuli in a different way for each of the 8 *S*s.

RESULTS

Differences among classifications.—Although all six classifications were constructed by dividing the same group of eight colors into two equal subgroups, there was a three-fold variation in difficulty from Classification A, at the easiest extreme, to Classifications E and F, at the hardest. Specifically, the mean numbers of errors (averaged over all *S*s and all positions within the sequence of four problems administered to each *S*) were, in alphabetical order from A to F, 18.8, 26.7, 42.3, 38.1, 57.6, and 56.5. Moreover, this same overall trend was evident within each of the four positions of administration taken separately. However, since the results for the four separate positions are not strictly independent, a test of the reliability of these differences was confined to the first position of administration only. Whether or not the error scores for this first position were logarithmically transformed (as in the earlier study by Shepard et al., 1961), these differences among classifications proved to be highly significant (e.g., for the log scores, $F = 8.37$, $p < .001$).

Effect of prior experience on other classifications.—In addition to the differences among the classifications per se, there was also a tendency to make fewer errors when classifications were administered late rather than early in the series of four positions. In particular, the mean numbers of errors (averaged over all six classifications) for the first through fourth positions were, in that order, 50.3, 43.0, 39.4, and 27.4. However, the absence of an apparent interaction between classifications and positions suggests that this gradual decline in errors over the four positions resulted more from a general increase in familiarity with the stimuli and with the paired-associate task than from a specific transfer from one classification to the next. (This, of course, is what we had hoped to achieve by choosing all six classifications to be "orthogonal.") Unfortunately, a conclusive statistical test of the interaction in question is precluded in the present experimental design. Nevertheless, in view of the fact that the classifications exhibited the same general trend in difficulty within both the early and late positions, it seemed reasonable to average the data for each classification across all four positions in order to permit a more sensitive test of the predictions to be considered next.

Calculation of the predicted difficulties.—The predictions were based upon the matrix obtained in the previous study of identification learning (Shepard, 1958, p. 513). For each pair of stimuli, this matrix gives the estimated mean probability with which the one stimulus was mistaken for the other over the series of 200 presentations. Since S_4 was not included among the present stimuli, however, the fourth row and fourth column of the original 9×9 matrix were deleted. Every entry was then divided by the new sum for its row in order to reinstate the required condition that the probabilities in each row sum to unity. Such a procedure is justified in part by Clarke's "constant-ratio rule for confusion matrices" (Clarke, 1957; Luce, 1959, p. 12). All 64 entries were then multiplied by 200/8 to convert them into the total numbers of confusions to be expected during all 200 trials. Apart from two corrections shortly to be mentioned, the predicted number of errors was obtained separately for each of the eight stimuli in each of the six classifications simply by summing the appropriate four entries in the row of the matrix corresponding to that stimulus. The total number of errors predicted for any classification as a whole, then, was furnished immediately by the sum of the predicted numbers of errors for all eight stimuli.

Actually, as already indicated, two corrections were applied prior to making these calculations. One was to compensate for an apparent overall difference in learning ability between the old and the new population of *S*s. The other was to compensate for the tendency of *S*s to make fewer errors on classifications given towards the end of the series of four. By coincidence, though, these two corrections turned out to have opposite and essentially equal effects (one multiplied all predicted values by 1.31 while the other multiplied all obtained values by 1.26). Therefore the application of the corrections had a negligible effect on the accuracy of the prediction.[2]

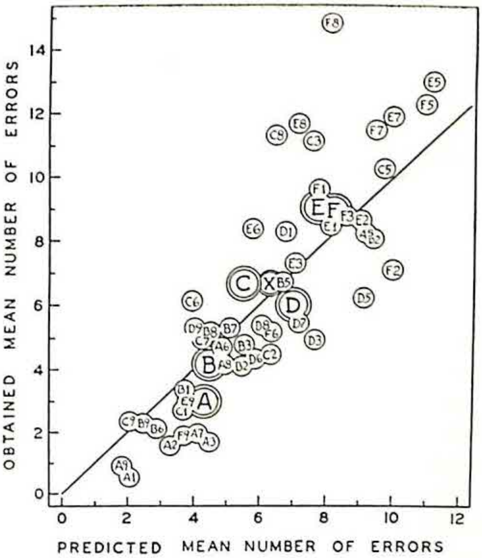


FIG. 3. Obtained vs. predicted difficulties. (The mean numbers of errors per stimulus per *S* are plotted [for each classification and for each stimulus within each classification] against the corresponding numbers predicted from the earlier identification experiment.)

Accuracy of the prediction.—Figure 3 shows the results actually obtained during classification learning plotted against the corresponding predictions calculated from the results of the earlier experiment on identification learning. For example, the center of the lowest circle (labeled A1) relates the (corrected) obtained to the (corrected) predicted mean number of errors (per *S*) on Stimulus S_1 in Classification A. Each of the six larger double-circled points (A through F) gives the mean number of errors per stimulus for a classification as a whole. The double-circled point labeled X (in the center of the plot) is based upon data from 4 *S*s of a preliminary experiment using another classification of these same stimuli.

---

[2] The need for the first correction was indicated by the fact that the mean number of errors per *S* during identification learning (i.e., the sum of all 56 off-diagonal entries in the 8 × 8 matrix) was 68.9 for the 36 graduate students in the earlier experiment as opposed to 90.4 for the 8 clerical employees in the present identification condition. It therefore seemed appropriate to adjust all entries in the matrix obtained from the previous 36 *S*s by multiplying them by the constant 90.4/68.9 = 1.31. (The decision to use a multiplicative correction, here, was based on the observation that a plot of the off-diagonal entries for the old vs. the new *S*s did not indicate any systematic departure from the straight line through the origin with slope 1.31.)

The need for the second correction was indicated by the fact that, whereas the mean number of errors for classifications in all four positions of administration was 40.0, the mean for classifications in only the first position was 50.3. Therefore, in order to make the classification data comparable with the earlier identification data (which were all obtained during first learning), it seemed desirable to multiply the classification data (averaged over all four positions) by the constant 50.3/40.0 = 1.26. As noted above, since the ratio of the two corrections is near unity (1.04), they effectively canceled each other.

Figure 3 reveals a rather strong relation between the predicted and obtained numbers of errors. The question remains, however, as to whether the residual scatter of the points about the prediction line represents a flaw in the predictive machinery or simply chance fluctuations in the relatively small sample of empirical data. In order to answer this question it was necessary to secure estimates of the variabilities of both the obtained and the predicted numbers. This was accomplished by dividing both groups of *S*s into two independent subgroups. First, that is, the 36 *S*s from the earlier identification learning experiment were divided at random into two groups of 18 *S*s each and, then, the predicted numbers for the various classifications were derived in exactly the same manner as before, but from each subgroup independently. Second, the 30 *S*s from the present classification learning experiment were divided into two groups of 15 *S*s each in a manner that was random except for the constraint that, in each of the two subgroups, each of the six classifications was represented exactly 10 times (and approximately equally often from each of the four positions of administration). In this way four independent 8 × 6 matrices were obtained each of which specified a mean number of errors for each of the eight stimuli in each of the six classifications. Again, in order to reduce the correlation between the variance and the mean for a group of scores, all entries in these matrices were logarithmically transformed. The two predicted matrices were then paired at random with the two obtained matrices and the differences between the corresponding entries in the paired matrices were computed.

The two resulting 8 × 6 matrices of differences were then treated as two "pseudo replications" or the random factor in a "mixed" three-way design and the corresponding analysis of variance was applied (e.g., see Edwards, 1960, pp. 307–309, where his "subjects" factor would correspond to our "pseudo replications" factor). Although this analysis is not strictly correct, here, because each of the present *S*s learned only four of the six classifications, the extensive counterbalancing is thought to be sufficient to prevent any serious bias.

If the scatter of the points about the prediction line in Fig. 3 resulted only from sampling fluctuations, then, this analysis should fail to detect any significant influence (upon these differences) of either the classifications, the stimuli, or their interaction. On the contrary, however, when all interactions of the random (or pseudo replications) factor with the other factors were pooled as the error term (with 47 *df*), significant differences among these discrepancies were found to be attributable both to classifications ($F = 6.82$, $p < .01$) and to stimuli ($F = 3.78$, $p < .01$). The interaction, however, was not significant ($F = 1.37$). Thus it appears that the departures of the obtained from the predicted numbers are not merely chance fluctuations. Moreover, these significant effects cannot be explained on the assumption of either systematic or unsystematic errors in the corrections applied to the obtained and predicted numbers. In fact, these multiplicative corrections were in effect converted, by the log transformation, into additive corrections and, hence, they essentially affected only that 1 *df* (out of the total 96) that was excluded from the analysis of variance of the differences. In conclusion, although the prediction accounted for a substantial fraction of the total variance of the obtained data (63%), the residual 37% is probably attributable at least in part to reliable departures from prediction.

DISCUSSION

An instructive comparison can now be made between the prediction attempted here and the similar one attempted in the earlier study by Shepard, Hovland, and Jenkins (1961). In order to facilitate comparison, we shall consider only the prediction of the total number of errors (per *S*) for each of the six classifications as a whole. (In the earlier study predictions were not computed for each stimulus individually.) As a rough measure of the accuracy of a prediction, here, we again use the fraction of the total variance of the obtained numbers that is accounted for by the predicted numbers (i. e., $1 - \Sigma(y - x)^2/\Sigma y^2$, where x and y denote the deviations of the predicted and the obtained numbers, respectively, from their means). With respect to the total number of errors made on each classification of the Munsell colors, then, 78% of the variance was accounted for by the prediction in the present experiment. In the earlier experiment with highly analyzable stimuli, by contrast, only 41% of the variance was accounted for by a prediction of essentially the same kind. This difference in the accuracies of the two predictions can also be seen quite directly by comparing the fit of the six points A through F in the present Fig. 3 with the fit of the six points I through VI in Fig. 13 of the earlier report (Shepard et al., 1961, p. 27). Furthermore, the superiority of the present prediction cannot be accounted for by a difference in the total range of difficulty covered by the classifications in the two experiments, for this range was actually greater in the earlier experiment (though the total variances of the obtained numbers of errors were nearly the same for the two studies). Thus the substantial improvement in the predictions from our explicit interpretation of the principle of stimulus generalization is most likely a consequence of the present substitution of the color stimuli for the relatively much more analyzable stimuli used in the earlier experiment.

According to arguments developed in that earlier study, the relative failure of the prediction there was a consequence of a second process, distinct from pure stimulus generalization. When stimuli are readily analyzed into perceptually isolated dimensions and values, it was argued, *S*s tend to narrow their attention to only those dimensions that are relevant for the given classification. The extent to which they generalize across the unattended dimensions then drops below the level that would be predicted from identification learning (where all dimensions are relevant). Now the present stimuli are clearly less analyzable, in this sense, than the stimuli used in the earlier experiment. For example, *S*s immediately say of two triangles—one large and black, the other small and white—that they differ in two respects (namely, size and color). But, when *S*s are shown two colors from the present experiment, they typically report simply that the two colors are "different shades of red," or that one looks like "tomato soup" while the other is more "rose colored." If such stimuli are not analyzed into components, then, the errors during classification learning should be largely accounted for in terms of the pair-wise confusions between stimuli as determined from identification learning. The relative success of the prediction with the color stimuli is therefore construed as lending further support to the arguments already put forward, tentatively, in the earlier report.

Since the deviations of the points from the prediction line in Fig. 3 are evidently not exclusively attributable to statistical fluctuations, however, the tested interpretation of the principle of generalization apparently does not capture all that is going on even when the stimuli are restricted to colors. The most obvious pattern in the present departures from prediction is that the overall slope of the obtained points seems greater than predicted. Interestingly, this is the most striking discrepancy that emerged in two earlier attempts to predict from the same generalization model (Shepard, 1961; Shepard et al., 1961, p. 27). This suggests that, over and above pure stimulus generalization, a second general process (perhaps something like abstraction) entered into all three of these studies—but, as anticipated, to a drastically reduced extent with the present color stimuli. The plausibility of this conjecture may be increased by noting, in Fig. 2, that the easy classifications tend to differ from the difficult ones in that their points can be roughly partitioned into the two subclasses simply by drawing a straight line through the two-dimensional space. Conceivably, under this condition the space becomes somehow "polarized" and, hence, renders the relevant direction (or dimension) through the space more salient. Much as in abstraction, then, the task would reduce to learning a division of the stimuli into two clusters along that single dimension.

REFERENCES

BAUM, M. H. Simple concept learning as a function of intralist generalization. *J. exp. Psychol.*, 1954, **47**, 89–94.

BUSS, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–503.

CLARKE, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29**, 715–720.

EDWARDS, A. L. *Experimental design in psychological research.* (2nd ed.) New York: Rinehart, 1960.

FRENCH, R. S. Number of common elements and consistency of reinforcement in a discrimination learning task. *J. exp. Psychol.*, 1953, **45**, 25–34.

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

HEIDBREDER, E. The attainment of concepts: III. The process. *J. Psychol.*, 1947, **24**, 93–138.

LUCE, R. D. *Individual choice behavior.* New York: Wiley, 1959.

MUNSELL COLOR COMPANY, *Munsell book of color.* (Abridged ed.) Baltimore: MCC, 1945.

NEWMAN, S. E. Effects of contiguity and similarity on the learning of concepts. *J. exp. Psychol.*, 1956, **52**, 349–353.

OSEAS, L., & UNDERWOOD, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43**, 143–148.

SHEPARD, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 1957, **22**, 325–345.

SHEPARD, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *J. exp. Psychol.*, 1958, **55**, 509–523.

SHEPARD, R. N. Role of generalization in stimulus-response compatibility. *Percept. mot. Skills*, 1961, **13**, 59–62.

SHEPARD, R. N., HOVLAND, C. I., & JENKINS, H. M. Learning and memorization of classifications. *Psychol. Monogr.*, 1961, **75**(13, Whole No. 517).

(Received February 12, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 103–105

EAR PREFERENCE IN AUDITORY PERCEPTION[1]

M. P. BRYDEN

McGill University

In a dichotic listening experiment, normal adult Ss identified numbers presented to the right ear more accurately than numbers presented to the left ear and preferred to report the material from the right ear first. To evaluate the effect of order of report a further experiment was performed ($N = 32$) in which Ss were instructed to report each channel first equally often. When 3 or 4 digits were presented to each ear, most Ss identified material from the right ear more accurately than material from the left ear. When 5 digits were presented to each ear, this difference was not observed. The findings suggest that the auditory system is better organized for the perception of verbal material presented to the right ear.

In a recent paper, Kimura (1961) reported that both brain damaged and normal *S*s are more accurate in reporting material presented to the right ear in a dichotic listening experiment (Broadbent, 1954). Using the same procedure, the present author has obtained data from 92 normal *S*s. Fifty-four of these were more accurate in identifying material from the right ear, while only 20 were better on the left ear ($p < .001$, binomial test).

Most *S*s tended to report the material from the right ear before giving that from the left. Since Broadbent (1957) has shown that *S*s are more accurate in identifying material from the first channel reported, it seemed possible that the right-ear superiority resulted from this preference for reporting the material from the right ear first. To test this hypothesis, a product-moment correlation between the proportion of trials on which *S* began on the right ear was calculated. Omitting *S*s who identified all the material correctly on all the trials, the correlation was .48 ($N = 79$, $p < .001$).

The above finding indicates that one reason for the superior accuracy in reporting material presented to the right ear is that *S*s forget the material presented to the left ear before they have a chance to report it. It should be possible to eliminate this factor by controlling the order of report so that each channel is reported first equally often; any remaining right-ear superiority could then be attributed to a true perceptual difference. An experiment was therefore carried out to accomplish this.

METHOD

Thirty sequences, using the numbers from 1 to 10, were prepared in such a way that when played on a stereophonic tape recorder (Philips Model 3536A) through earphones (Koss Model SP-3), one number was heard at the left ear at the same time a second, different number was heard at the right ear. There were 10 sequences of three pairs, 10 of four pairs, and 10 of five pairs; all were presented at the rate of two pairs per sec.

A total of 32 college undergraduates were used as *S*s. None had any hearing impairment of which they were aware. In order to ensure that identification differences were a function of the *S*s and not of the recording, half the *S*s heard Channel 1 of the recording in their left ear and half heard it in their right ear. In addition, the order of presenting the material and of instructing the order of report were counterbalanced.

---

[1] The research of this paper was supported by the Defence Research Board of Canada, Grant No. 9401-11.

All *S*s were instructed to report all the material they heard in one ear before reporting any they heard in the other ear. Half the time it was the left ear they reported first and half the time the right ear: each of the 30 digit sequences was heard once under each of these conditions. In scoring the data, *S*s were credited with a correct response whenever they identified a number that had actually been presented, whether or not they were successful in following the ordering instructions. However, a detailed record of errors in order of report was kept.

RESULTS

Table 1 shows the mean number of digits identified under each condition for the two orders of report combined. Each *S* can be classified as to whether he was more accurate in reporting material from the left ear or from the right ear. These data are also shown in Table 1. In the three- and four-pair conditions, a binomial expansion reveals that significantly more *S*s were superior in identifying material presented to the right ear than would be expected by chance. Such a difference does not appear with five pairs of numbers.

The *S*s were not entirely successful in using the instructed order of report, particularly when there were five pairs of numbers. Table 2 shows the frequency with which *S*s reported one or more numbers from the second ear before reporting all the numbers they eventually reported from the first ear.

TABLE 1

COMPARISON BETWEEN ACCURACY OF IDENTIFICATION FOR LEFT AND RIGHT EARS

Cond.	Digits Correctly Identified				No. *S*s Better on	
	Left Ear		Right Ear			
	Mean	*SD*	Mean	*SD*	Left	Right
Three-pair	55.25	3.51	56.09	4.99	8	20*
Four-pair	61.81	10.55	65.22	10.79	8	23*
Five-pair	63.97	9.48	64.63	11.90	15	17

* Left-right difference; $p < .05$; two-tailed binomial test.

TABLE 2

FREQUENCY OF ERRORS IN ORDER OF REPORT

Cond.	Mean Errors		
	Left Ear First	Right Ear First	Total
Three-pair	1.19	1.03	2.22
Four-pair	2.16	1.66	3.81
Five-pair	4.13	3.89	8.03
Total	7.48	6.58	14.06

There is a slight tendency for *S*s to make more errors in order of report when instructed to give the material from the left ear first. To ensure that this increased frequency of errors was not responsible for the right ear superiority demonstrated in the three- and four-pair conditions, the data were reanalyzed to include only those trials on which no error in order of report was made. The mean difference between left- and right-ear scores was calculated for those errorless trials on which *S* reported the left ear first, and for the trials on which he reported the right ear first. These two figures were then averaged, for each *S* separately, to give a measure of right-ear superiority unbiased by order of report. This measure is positive when *S* is more accurate on the right, and negative when he is better on the left. The mean value of this difference score is shown for the various conditions in Table 3. A similar score for all trials, whether or not errors of order were made, was determined by calculating the left-right difference per trial. This is also shown in Table 3, as are comparable figures from an earlier free recall experiment.

Even when only the trials in which no error was made in order of report are used, *S*s tend to be more accurate in identifying material presented to the right ear. With four pairs of numbers, 23 *S*s are better on the right and 9 better on the left ($p = .02$); this is about the same proportion of *S*s that are better on the right under free recall conditions. With three pairs, 16 *S*s are better on the right, 8 on the left, and 8 show no difference. While this does not reach statistical significance ($p = .15$), the trend is in the expected direction and so few errors of identification are made that some of the left-right differences may be fortuitous. Finally, there is no difference in accuracy between left and right ear for the five-pair condition: 16 *S*s are better on the left and 16 better on the right. When five pairs of numbers are presented, *S* has more material than he can handle. In this condition, a great many errors in order of report occur. Even when the order is correct, *S* often switches from one channel to the other after giving only one or two digits, or alternatively, reports material from only one channel. The failure to observe any left-right difference under this condition might well stem from the confusion produced by such a large amount of material.

The data of Table 3 also indicate that left-right differences obtained by instructing order of report are much smaller than those obtained with free recall. This would be expected, if there were a bias produced by more frequently reporting the right ear first.

TABLE 3

MEAN DIFFERENCES IN ACCURACY BETWEEN RIGHT AND LEFT EARS IN DIFFERENT CONDITIONS

Order of Report	Cond.		
	Three-Pair	Four-Pair	Five-Pair
Free recall	.187*	.417*	.875
Instructed, all trials	.042*	.171*	.033
Instructed, errorless trials only	.018	.151*	−.098

* More *S*s are superior on the right; $p < .05$; two-tailed binomial test.

DISCUSSION

The data presented in this paper demonstrate several points. First of all, the right-ear dominance in dichotic listening tasks reported by Kimura (1961) has been verified with a sizeable group of normal *S*s. Secondly, this right-ear superiority has been shown to be related to a preference for reporting the material presented to the right ear first: to some extent, accuracy in identifying material presented to the left ear is inferior because there has been more of an opportunity to forget this material. Finally, right-ear dominance has been shown to exist even when order of report is controlled.

The interpretation offered by Kimura (1961) would seem appropriate to these data. It seems plausible to assume that the right-ear superiority observed when order of report is controlled is due to a true perceptual difference, perhaps specific to verbal material. According to this hypothesis, material presented to the right ear would more actively excite the appropriate trace systems in the cortex. As a result, the traces of the material presented to the right ear would take longer to fade below response threshold, and thus could be more accurately reported toward the end of the response sequence than could material presented to the left ear. Furthermore, such differences in trace activity could also provide a basis for reporting the right channel first; *S* may select the stronger of the two alternatives to begin reporting with.

REFERENCES

BROADBENT, D. E. The role of auditory localization in attention and memory span. *J. exp. Psychol.*, 1954, **47,** 191–196.

BROADBENT, D. E. Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, 1957, **9,** 1–11.

KIMURA, D. Cerebral dominance and the perception of verbal stimuli. *Canad. J. Psychol.*, 1961, **15,** 166–171.

(Received February 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 106–107

SUPPLEMENTARY REPORTS

EFFECTS OF PROBABILITY OF REWARD AND SPEED REQUIREMENT ON HUMAN PERFORMANCE[1]

SHEILA G. ZIPF

University of California, Berkeley

An experiment was conducted to investigate 6 levels of probability of reward for success (PrR) and 6 levels of required increase in speed (Req) as they effect change in speed of performance of a simple motor task. The index of performance used is a ratio: improvement in speed relative to possible improvement. Both variables have a significant effect on performance. The PrR has a linear effect while Req has both linear and quadratic components. It is suggested that the effect of Req is similar to the increasing then decreasing amount of performance found with increasing activation.

Zipf (1961) has reported a positive relation of required speed (Req) and performance and a possible positive relation of probability of reward (PrR) to performance. In the present experiment, more levels of each of these variables were examined. The six levels of Req utilized were: 1.0, 1.2, 1.4, 1.6, 1.8, and 2.0 times the practice level. The six levels of PrR were: 0 (no mention of a reward), .2, .4, .6, .8, and 1.0 chance of receiving $2 if the required speed were achieved.

Method.—The general procedure was similar to that of the previous study. The task, however, was tapping three holes per circle on a page with 100 circles. See the tapping test in MacQuarrie (1927). A different sheet was used for each trial. Trials were 10 sec. long with a 15-sec. intertrial interval. Rest periods were 5 min. long. The *S*s, 108 men and women from introductory classes in psychology at the University of California, were tested individually and were instructed on the meaning of probability.

The order of events was as follows: Trials 1–40, rest; Trials 41–45, rest during which Req and PrR instructions were given; Trial 46, rest during which each *S* received $2 and was asked to go as fast as possible on the next trials. An incentive of $10 was offered for the fastest performance. Trials 47–51 completed the series. The Req and PrR instructions were similar to those of the earlier experiment except that no mention of *S*'s probability of success was made and all incentives were $2. A Req of 2.0, for example was described as follows. "On the last 5 trials, your average was about *15* circles. On the next trial, which will also be 10 sec. long, I would like you to go two times as fast as that. That is, I would like you to make three taps in at least *30* circles on the next trial." Italicized items were filled in by *E*.

Trial 41 was used as the measure of base rate T_0, Trial 46 is the test trial, T_t and the fastest of Trials 46–51 was taken as the upper limit performance T_u. The index of performance used was $T_t - T_0/T_u - T_0$, or improvement in speed following Req and PrR instructions relative to possible improvement. Neither T_0 nor T_u were significantly related to the experimental conditions.

Results.—To correct a significant nonhomogeneity of variance, the performance index was transformed to $\text{Log}(X + 2)$. The analysis of variance on these scores is summarized in Table 1. The required speed of performance is highly significant and has both linear and quadratic components. The anticipated linear effect of PrR just reaches significance. The two variables do not interact.

The quadratic effect of Req is partly due to performance at Req of 2.0 being less than at Req of 1.8. Both of these requirements were beyond the upper limit performance of all but one of the *S*s. The highest Req may be "too motivating." The result is similar to the finding of Stennett (1957) who found a U shaped curve of human motor performance with increasing activation. Activation was produced by instructions to improve performance and offering of incentives.

---

[1] This investigation was supported by a postdoctoral ellowship (MF-9450) from the National Institute of Mental Health, Public Health Service and was conucted at the University of California, Berkeley. The writer wishes to express appreciation to Richard Crutchfield who sponsored the research and to the Department of Psychology for the facilities provided.

TABLE 1

Analyses of Variance of Transformed Performance Scores

Source	*df*	*MS*	*F*
Requirement (Req)	(5)		
Linear	1	.052	13.00**
Quadratic	1	.085	21.25**
Residual	3	.006	1.40
Probability of Reward (PrR)	(5)		
Linear	1	.016	4.00*
Residual	4	.004	1.00
Req × PrR	25	.004	1.00
Error	72	.004	

Note.—Log $\frac{T_t - T_0}{T_u - T_0} + 2$.

* $p < .05$.
** $p < .01$.

The major deviation from linearity of the PrR effect is the relatively high performance of the *S*s who were told nothing about a reward. Perhaps no reward should not be placed on the same dimension as PrR.

The performance index $T_t - T_0/T_u - T_0$ had been selected prior to collecting data as a method of controlling individual differences in ability and initial motivation on the task. The analysis of Table 1 was also performed on T_t and on the ratio T_t/T_0 which was used in the earlier experiment. The linear component of PrR is not significant with either of these indices. The quadratic component of Req is significant with both but the linear component is significant only with T_t. The new index appears to be a more sensitive measure than either the absolute score or the simple ratio score.

References

MacQuarrie, T. W. A mechanical ability test. *J. Personnel Res.*, 1927, **5**, 329–337.

Stennett, R. G. The relationship of performance level to level of arousal. *J. exp. Psychol.*, 1957, **54**, 54–61.

Zipf, S. G. The effects of amount of reward, requirement, and several related probabilities on human performance. *J. exp. Psychol.*, 1961, **62**, 503–509.

(Received January 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 107–108

ON MOTION PARALLAX AND PERCEIVED DEPTH[1]

OLIN W. SMITH and PATRICIA CAIN SMITH

Cornell University

The adequacy of the combination of a "two-velocity" motion parallax and a continuous gradient of horizontal velocities as a stimulus for perceived depth and distance between 2 surfaces was tested on 182 children and adults. A shadow caster was used. Order of the surfaces in depth was perceived as would be predicted by motion parallax by all Ss except one. The distance of separation in depth of the surfaces was indeterminate. The perceived slant of the surfaces was not as predicted by parallax. No developmental differences were observed.

Gibson, Gibson, Smith, and Flock (1959) tested a "two-velocity" motion parallax (Type 1), and a graded flow of velocities (Type 2) as stimuli for perceived depths and distances. For both types, an eye and a point light source were placed symmetrically at equal distances on either side of a translucent screen. For Type 1, two plastic panels, dusted with talcum powder, were placed parallel to the screen at different distances between light and screen. Movement of the panels parallel to the screen produced two sets of shadows, each having a different translatory velocity. Type 2 was generated by substituting for the dual panels one paint-spattered panel tilted toward the screen at an angle of 45°. The carriage for this panel moved parallel to the screen, thus producing a continuous gradient of velocities increasing from top to bottom of the screen. Depth was perceived by all 46 *S*s on Type 1, and by 45 of 49 *S*s on Type 2. Two surfaces were perceived as parallel by all *S*s on Type 1 while the 45 *S*s on Type 2 who perceived depth saw the surface as slanted. The absolute magnitude of the perceived depths was indeterminate for both types.

The perceived order in depth of the two sets of shadows of Type 1 was not consistent within or between *S*s; moreover, not all *S*s perceived the surface of Type 2 as slanted. A combination of the two types of stimuli using two slanted, moving surfaces was predicted to remove these ambiguities since: (*a*) all *S*s perceived the surfaces of Type 1 as separated in depth, while (*b*) no reversals

---

[1] This research was supported in part by Public Health Service Grant No. B-1586 (C2) to Cornell University by the National Institute of Neurological Diseases and Blindness.

of slant of the surface of Type 2 were noted. A test of the above prediction is reported here. Children as well as adults were tested in order to assess possible differences in perception attributable to differences in age.

Method.—On a transparent screen, opaque paper framed an opening which subtended to the eyes of *S* at 48 in. horizontal and vertical angles of 46°4′ and 28°4′, respectively. A 12–16 v., dc, 32-cp, automobile bulb was placed behind the screen at 48 in., symmetrically with the eyes. From *S*'s position, overall illumination of the aperture was about 5 ft-c (slightly greater in the center). Two parallel plastic panels, 36 in. long and 12 in. high, each dusted lightly on one side with talcum powder, were mounted 8 in. apart on a carriage which moved parallel to the screen. When the carriage was centered, the line from light to eyes bisected the panels vertically and horizontally. The point midway between the panels was 24 in. from the light. The eye-light line made a horizontal angle of 79° with both panels.

The carriage traveled to and fro about 17 in. in each direction at an angular rate of circa 2°30′/sec. Two sets of shadows were cast, each of which included a different graded flow of horizontal velocities, thus constituting a combination of Types 1 and 2. The *S* observed with unshielded binocular vision in a darkened room. The light source was activated only after the carriage was set in motion.

Preliminary observations showed that: (*a*) static panels induced perceptions of a flat, depthless surface, and (*b*) young children had difficulty in specifying which appeared in front even when the panels were in motion. An equilateral triangle, $1\frac{5}{8}$ in. on a side, was therefore fixed on the center of the panel closest to the screen. Two paper fish, $1\frac{1}{2}$ in. long and 2 in. wide (chosen for interest to children), were attached to the other panel. One was located horizontally in the middle and 2 in. from the top. The second was displaced $1\frac{1}{2}$ in. from the horizontal midpoint, 5 in. below the top.

The *S*s were drawn from the Cornell community: 97 children evenly distributed in age (60–143 mo.) and 85 "adults" (12–42 yr. of age). Each *S* was first instructed to describe what he saw. If he did not comment on the order in apparent depth of the triangle and fish, he was asked if they appeared to be at the same distance from him. Given a response that one appeared in front, *E* instructed *S* to designate on a yardstick the extent of the apparent separation.

Results and discussion.—Most children and adults reported voluntarily that the fish appeared to be in front of the triangle. Upon questioning, 96 out of 97 children and all adults gave this report. The combination of the two types of parallax removed the spontaneous reversals of the front-back relationships of the two surfaces as reported previously for Type 1 parallax.

The apparent distance of separation of the surfaces was indeterminate, as in previous studies. The *S*s were forced to guess, resulting in judgments with small means about equal to the *SD*s.

No differences in perceptions attributable to differences in age were detected except that children were apt to report the fish as swimming.

The flow of velocities across the horizontal extent of the screen was ineffectual as a stimulus for perceptions of a slanted surface. These results are incompatible with those reported above for Type 2 parallax and indicate that the continuous gradient of motions proposed by Gibson, Olum, and Rosenblatt (1955) as a stimulus for perceived slant-depth lacks generality and may be situationally specific.

Motion parallax is a generic term subsuming all instances of parallactic motions. At present, published taxonomies do not exist for: (*a*) different types of these motions, (*b*) physical situations which cause these motions, and (*c*) perceptions for which combinations of *a* and *b* may be sufficient. Consequently, it seems unlikely that the findings of any one experiment on a particular type of motion can possibly be conclusive for all instances of parallax.

The results of this experiment in conjunction with those of Gibson, Gibson, Smith, and Flock (1959), and Smith and Smith (1961) demonstrate that the only general statement which holds for both perspectivities and perspective transformations containing parallactic motions (see Adler, 1958) is that they may be stimuli sufficient for perceived motions. Motion parallax without further specification does not insure perceptions of: (*a*) depth, (*b*) distance, (*c*) order of surfaces in depth, or (*d*) the slant or curvature of surfaces.

REFERENCES

ADLER, C. F. *Modern geometry*. New York: McGraw-Hill, 1958.

GIBSON, E. J., GIBSON, J. J., SMITH, O. W., & FLOCK, H. Motion parallax as a determinant of perceived depth. *J. exp. Psychol.*, 1959, **58**, 40–51.

GIBSON, J. J., OLUM, P., & ROSENBLATT, F. Parallax and perspective during aircraft landings. *Amer. J. Psychol.*, 1955, **68**, 372–385.

SMITH, P. C., & SMITH, O. W. Veridical perceptions of cylindricality: A problem of depth discrimination and object identification. *J. exp. Psychol.*, 1961, **62**, 145–152.

(Received February 14, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 1, 109–112

HUMAN HEART RATE RESPONSES DURING EXPERIMENTALLY INDUCED ANXIETY: A FOLLOW-UP

R. STEPHEN JENKS[1] AND GEORGE E. DEANE[2]

College of Wooster

Base-level measurements of cardiac activity were made while Ss watched the sequence of Numbers 1 through 12 appear on a memory drum. Half of the Ss were then told to expect shock and half told to expect an extremely loud tone during Number 8, 9, or 10. Half of the shock Ss received 1 shock and half of the tone Ss received 1 tone. Other Ss received no noxious stimulus. All Ss showed an acceleration in rate during Numbers 1 through 6 and a deceleration during Numbers 8 through 10. The amplitude of the acceleration or deceleration was not dependent upon whether or not the noxious stimulus was received, but the shock-anticipation group showed greater acceleration than the tone group.

In an earlier study Deane (1961) showed that when *S*s were told to expect shock during the number 10 in a sequence of visually presented numbers from 1 to 12 the heart rate accelerated early in the number series and decelerated at the point where shock was expected. The present study is designed to answer the following questions: (*a*) Are the acceleration and deceleration noted in the previous study specific to the use of shock as the noxious stimulus or will a loud tone have the same effect? (*b*) Is the deceleration a very brief effect which might be associated with a short respiratory gasp just before the noxious stimulus is expected, or can the period of deceleration be extended by informing the *S* to expect the noxious stimulus over a more extended period of time?

Method.—The *S*s were 60 male students at the College of Wooster. A Grass Model 5 polygraph was used to record heart rate and to indicate the sequence of numbers presented on the memory drum at the rate of one every 3 sec. The shock was of 1-sec. duration and was supplied by an Applegate Model 288 constant current stimulator set to deliver 1.5 ma. The 1-sec. tone of 1600 cps which was delivered through a pair of WROL Model N-130 crystal headphones was generated by a Heathkit Model AG-9A signal generator set for maximum output.

Base-level measurements were made as in the earlier study and at the end of the tenth base-level presentation *E* gave *S* one of two sets of instructions. In Groups A and B each *S* was instructed that now sometime during one of the numbers 8, 9, or 10 he would receive an electric shock on his fingers. He was told that he might not receive a shock on every trial and, in fact, *S*s in Group A received only one shock during Number 10 on the first of six trials, and *S*s in Group B never received a shock. In Groups C and D each *S* was instructed that now sometime during one of the numbers 8, 9, or 10 he would hear an extremely loud and unpleasant tone through the headphones. He was told he might not receive the tone on every trial, and, in fact, *S*s in Group C received only one tone during the number 10 on the first of six trials, and *S*s in Group D never received a tone.

Results and discussion.—In Fig. 1 are plotted the results for each of the four groups. At Stimuli 1 → 6 is plotted the mean rate during presentation of these numbers. At each of the points 7 through 12 is plotted the mean rate of the middle three beats occurring during presentation of these numbers. In every case the solid line with the filled circles is a mean of the last three base-level presentations (B8–10), the solid line with the open circles is a mean of all six trials (T1–6) when a noxious stimulus was anticipated, and the solid line with the filled squares is a mean of the last five trials (T2–6) when a noxious stimulus was anticipated.

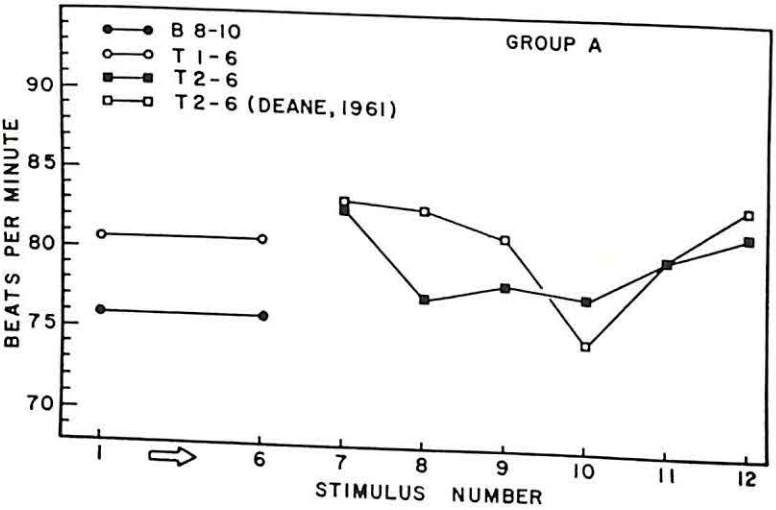
Figure 1 indicates that for all groups the effect of instructions regarding an impending noxious stimulus was to accelerate the rate over the base level during the first six stimulus numbers. An analysis of variance was performed on the mean cardiac rate of all *S*s during Stimuli 1 through 6 for the mean of Trials B8–10 and T1–6. Table 1 shows that the Trials effect is significant at better than the .005 level. Figure 1 also indicates that the amount of acceleration was greater for the groups given shock instructions than for the groups given tone instructions. This is supported by the Trials × Instructions interaction which is significant at better than the .05 level. In order to show that the tone instruction groups alone did exhibit a significant acceleration over the base level a t test was made between the mean rate during Trials B8–10 and the mean rate during Trials T1–6 for Groups C and D combined. This resulted in $t(29) = 4.17$; $p < .01$, two-tailed. The significant Instructions effect indicates that over all trials the groups given shock instructions had a higher mean rate.

---

[1] Now at Case Institute of Technology.

[2] Now at Harpur College.

Figure 1 also shows that for all groups there was a deceleration in rate on Trials T2–6 during Stimulus Numbers 8, 9, and 10 where the noxious stimulus was anticipated. An analysis of variance was performed on the mean difference scores (mean rate during Stimuli 4, 5, and 6 minus mean rate during Stimuli 8, 9, and 10) for Trials T2–6. Trial T1 was omitted from this analysis since two groups actually received the noxious stimulus on this trial, causing an additional disturbance in rate during Number 10. Table 1 shows that the Number effect is significant at better than the .005 level. The nonsignificant Number × Instructions interaction indicates that the amount of deceleration was as great for those *S*s given tone instructions as for those given shock instructions. The significant Instructions effect again indicates that



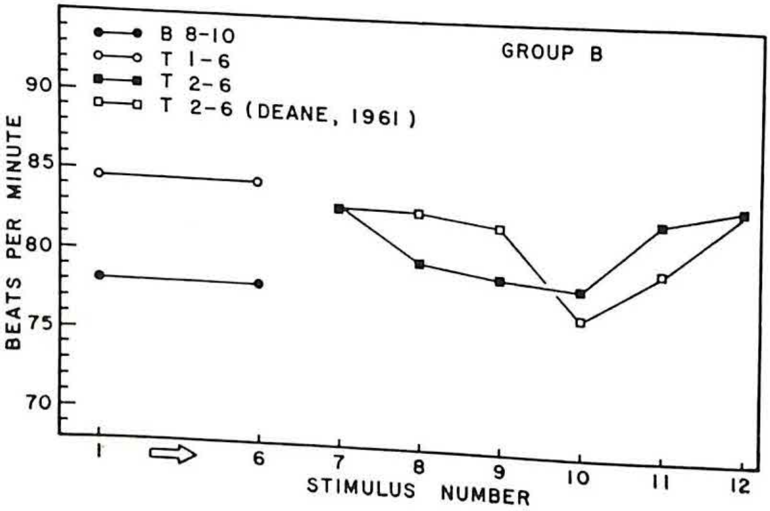


Figure 1—Continued

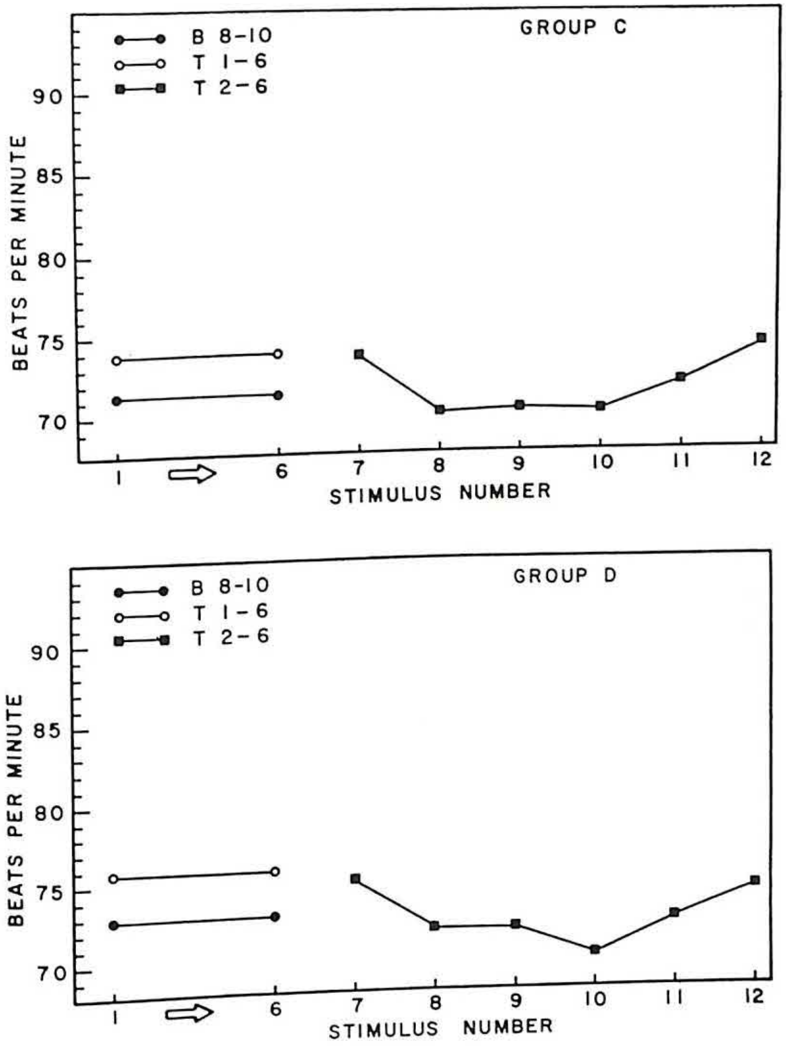


Fig. 1. Mean cardiac rate during the number stimuli with trials as the parameter.

TABLE 1

Analysis of Variance of Acceleration and Deceleration Effects

Acceleration Effect			Deceleration Effect		
Source	*df*	*F*	Source	*df*	*F*
Between			Between		
Instructions (I)	1	5.65*	I	1	8.72**
Stimulation (S)	1	0.84	S	1	0.42
I × S	1	0.07	I × S	1	0.02
Error (*MS*)	56	(21088.40)	Error (*MS*)	56	(21896.93)
Within			Within		
Trials (T)	1	41.39**	Numbers (N)	1	76.18**
T × I	1	5.12*	N × I	1	0.00
T × S	1	0.77	N × S	1	0.01
T × I × S	1	1.53	N × I × S	1	0.01
Error (*MS*)	56	(1186.88)	Error (*MS*)	56	(43475.66)

* $p < .05$.
** $p < .01$.

during all stimulus numbers the rate for the shock-instruction groups was greater than for the tone-instruction groups. None of the interactions was significant.

The present study provides further evidence that there are two opposing heart rate responses during experimentally induced anxiety and also indicates that these effects are not specific to the use of shock as the noxious stimulus. The lines with the open squares in Fig. 1 show the results of Groups A and B in the earlier study who were told to expect shock only at 10. In both studies the deceleration persisted during all numbers where shock was expected, or for a period from 3 sec. in the previous study to 9 sec. in the present one. Although Westcott and Huttenlocher (1961) showed that a respiratory gasp yields a consistent pattern of rapid acceleration followed by a deceleration reaching a minimum rate in about 6 sec., it would seem unlikely that this would account for the deceleration noted in both studies. Moreover, Deane and Zeaman (1958) have previously reported that neither respiratory rate, amplitude, nor percentage of time inhaling seems to be related to the acceleration and deceleration observed under the present conditions. The results of Groups B and D indicate again that verbal instructions regarding when to expect the noxious stimulus are sufficient to produce both cardiac effects without ever administering the noxious stimulus itself. The present results seem to lend support to the hypothesis that when *S* expects to receive a noxious stimulus a state of anxiety with its associated response of cardiac acceleration is aroused; and, in addition, if the *S* expects the noxious stimulus at a particular instant in time a state of fear with its associated response of cardiac deceleration is aroused immediately prior to and during the time the stimulus is expected.

REFERENCES

DEANE, G. E. Human heart rate responses during experimentally induced anxiety. *J. exp. Psychol.*, 1961, **61**, 489–493.

DEANE, G. E., & ZEAMAN, D. Human heart rate during anxiety. *Percept. mot. Skills*, 1958, **8**, 103–106.

WESCOTT, M. R., & HUTTENLOCHER, J. Cardiac conditioning: The effects and implications of controlled and uncontrolled respiration. *J. exp. Psychol.*, 1961, **61**, 353–359.

(Received February 26, 1962)

Journal of
Experimental Psychology

VOL. 65, NO. 2 FEBRUARY 1963

STIMULUS EXPOSURE TIME, BRIGHTNESS, AND SPATIAL FACTORS AS DETERMINANTS OF VISUAL PERCEPTION[1]

JAQUES KASWAN AND STEPHEN YOUNG

University of California, Los Angeles

In a forced choice task, 5 Ss chose between linear displays of dots arranged in pairs or evenly spaced presented at 8 exposure times ranging from .004 to .512 sec. and 8 intensities ranging from .09 to 11.84 mL. Intensity had little effect on accuracy. Accuracy was largely a function of exposure time and, for paired designs, the amount of relative spatial distance between to within pairs of dots. Detecting the presence or absence of stimuli was a joint function of exposure time and intensity and independent of relative spatial distance. The findings support the assumption that perception becomes differentiated over time, and indicate that relative spatial distance can be used as a stimulus measure of differentiation. A 2-phase process of the temporal development of perception is proposed.

Previous evidence suggests that when visual stimuli are presented at very brief exposure times (ET) they are seen as less differentiated than at longer ETs (Bender, 1938; Flavell & Dragnus, 1957; Kaswan, 1958). This effect has been interpreted to indicate that perception develops and becomes differentiated as a function of increasing ET—a process termed "microgenesis" by Flavell and Dragnus (1957).

Systematic investigation of this differentiation process requires the specification of the spatial stimulus parameters along which such differentiation may occur. Following previous work (Kaswan, 1958), variations in distance between dots arranged in a line are considered the relevant parameter in the present study. The specific measure of this parameter can be best illustrated with the stimuli used. When the distance between alternate dots is varied by a constant amount, pairs (Pr) of dots are generally perceived (see Pr designs, Table 1). The *difference* in distance between pairs *relative* to that within pairs is considered the spatial stimulus parameter relevant to perceptual differentiation. This parameter will be called relative spatial distance (RSD). Since the distance between pairs increases from Design Pr1 to Pr2 to Pr3 while the distance within pairs remains constant, these designs should reflect increasing facilitation of perceptual differentiation. In the evenly spaced (Ev) designs, the absolute distance between dots increases from Design Ev1 to Ev2 to Ev3, but the distance between dots within each design remains constant. Evenly spaced designs are therefore considered homogeneous with respect to variations in distance.

---

[1] This study was supported by a grant from the National Science Foundation (G-9589). Part of the data contained in this paper was presented at the 1961 American Psychological Association meetings. The authors would like to express their appreciation to George Boroczi for helping to run *Ss*. The generous availability of the facilities of the Western Data Processing Center was of substantial help in data analysis.

The assumption that less spatial differentiation is seen with decreasing ET implies that differences in spacing within a Pr design will be seen less frequently as ET is decreased. Thus, when *S*s are shown a Pr design with an Ev design as the alternative, the Pr design should be more often reported as Ev as ET is decreased. Since the perception of homogeneous spacing presumably occurs first in time, a high level of accuracy should be found for Ev designs even when ETs are brief. Thus, the difference in accuracy between Ev and Pr designs should be maximal at short ETs and should decrease as ET becomes longer. Further, the greater the relative spatial differences (RSD) between components in a design, the briefer should be the ET required to make it appear evenly spaced. Accordingly, it is assumed that the amount of time required for correct discrimination of a Pr design should increase in the order Pr3, Pr2, and Pr1.

It is conceivable that the proposed interaction of ET and RSD is not specifically a function of time but of the total energy of stimulation, which would involve as parameters both ET and degree of illumination. In fact, the evidence points to considerable reciprocity of intensity (I) and ET in the determination of perceptibility, as illustrated by the Bunsen-Roscoe law (e.g., Hunter & Sigler, 1940). This study is designed to explore the possibility of such reciprocity for stimuli whose components vary in relative spatial distance (RSD).

METHOD

Subjects.—Five paid male *S*s ranging from 21 to 29 yr. of age, were used in the experiment. The *S*s were UCLA seniors and graduate students and had some sophistication in psychology, but none had any conception of the purpose of the experiment. One *S* had minimal previous experience with tachistoscopic presentation. All *S*s had at least 20/30 uncorrected vision.

Apparatus.—An electronically timed tachistoscope was constructed for the experiment from a design similar to one used by Smith, Spence, and Klein (1959). Briefly, this instrument uses timed gating circuits and a logic system to provide for sequential presentation and independent timing of the fields. The apparatus was modified by the addition of a 1450-v. dc, 230-ma. full wave selenium power supply which was heavily loaded to provide 30 ma. of relatively pure direct current to each of the fields. Each field contained two vertically mounted 8-in. 4500°-K mercury argon lamps coated with magnesium tungstate and other short decay phosphors which were manufactured by the Aristo Corporation. With these improvements, it was found that a rectangular light pulse, with rise and decay times of approximately 250 μsec. could be obtained at exposure times as low as .004 sec. A switching system was added to allow rapid selection of the experimental exposure times. Continuous monitoring and calibration of the light output of the tachistoscope were obtained with a photomultiplier input to an oscilloscope. The photomultiplier was located directly below the eyepiece.

Two fields were used, a fixation field and a stimulus field. Semiopaque field stops in both fields provided an 8-in. square field with a discrete but minimally contrasting border. The two fields were matched for color and an unattenuated intensity of 11.84 mL. measured with a Spectra brightness spot meter (Photo Research Corporation). Intensity was attenuated by insertion of neutral density filters directly in front of the eyepiece, so that light output from the two fields was always identical.

Both the fixation and stimulus field were 40 in. from the eyepiece. A chin rest and close fitting face mask were used to insure minimal head movement. All viewing was binocular.

The fixation field contained two horizontal ¼-in. black dashes on a flat white surface located so that they would fall just inside the border and an equal distance on either side of the stimulus when the two fields were superimposed.

Stimuli.—Six stimuli were used (see Table 1). Each contained a single horizontal line of $\frac{1}{16}$-in. diameter black dots centrally located on flat white cards. The maximum visual angle subtended by a line of dots was 10° to either eye. Two basic linear arrays of dots, one evenly spaced and one paired, were used. Three paired patterns were obtained by successively doubling an increment in the distance between pairs. Corresponding evenly spaced patterns were obtained by spacing dots the average of the distance between and within pairs in the paired patterns. The dimensions of the stimuli are given in Table 1.

Procedure.—The experiment began with two practice sessions in which *S*s were shown only Designs Ev3 and Pr3 (Table 1). Each design was presented at eight different ETs at each of eight I's using the method of constant stimulus. The ETs were successively doubled from .004 sec. to .512 sec., thus ranging from -8 to -1 $\log_2$ sec. Intensities were successively halved from the maximum luminance of 11.84 to .093 mL., thus ranging from 3.5 to -3.5 $\log_2$ mL. In each session, eight I blocks of 16 exposures each (each ET for each design), were presented for a total of 128 exposures. The order of I blocks was randomized with the restraint that the initial block was always in the brighter half of the I range and that successive blocks were always within two ordinal I steps of each other. These restraints were intended to minimize dark adaptation problems. Design-ET combinations were randomized within each block with the restraint that each combination was used only once in each block of 16.

The *S*s were first shown either the Ev3 or Pr3 design and told to hit one of the two microswitch keys, located on either side of a finger rest, as quickly as possible whenever this design was presented. They were then shown the other design and instructed to respond with the other key. A warning signal was given approximately .5 sec. before each exposure with a manually operated buzzer in response to which *S* was asked to focus between the edge markers of the blank field. For 3 of the *S*s the key on the right side was associated with the Ev design, for 2 of the *S*s this key was associated with the Pr design. The designs were never verbally described by either *S*s or *E* until the end of the study. If *S* saw absolutely no indication of a design during an exposure (i.e., not even a blurred line), he was instructed to hit either key and then say the word "nothing" (NO). There was an interval of approximately 11 sec between exposures within a block of 16 exposures and 45 sec. between each block when I filters were changed.

TABLE 1

CRITICAL DIMENSIONS OF STIMULUS PATTERNS

Set	Pattern Designation	Number of Dots	Length of Line (Cm.)	Distance between Pairs	Cm.[a] within Pairs
1	Ev1	16	16.5		1.1
	Pr1	16	16.5	1.2	1.0
2	Ev2	16	17.7		1.2
	Pr2	16	17.7	1.4	1.0
3	Ev3	14	18.2		1.4
	Pr3	14	18.2	1.8	1.0

[a] For Ev patterns: distance between dots.

The experimental sessions were like the practice sessions. Each *S* was given one of the three Ev-Pr sets at each session, with the sequence of sets randomized, but all three sets were given before one was repeated. Each *S* had 43 sessions of 45 min. each, spaced over a 3-mo. period, with from three to five sessions per week. Each *S* thus received 13–15 replications of each I-ET-design combination, so that a total of almost 28,000 responses was recorded for the 5 *S*s combined.

RESULTS

RSD-ET relations.—The relevant results are shown largely in Fig. 1a and 2a. Percent correct values were obtained by dividing the number of correct responses by the combined number of correct and incorrect responses, excluding NO responses. There was a total of approximately 560 exposures for each point in Fig. 1 (a and b) and 2 (a and b) so that the *N* for each point can be computed by

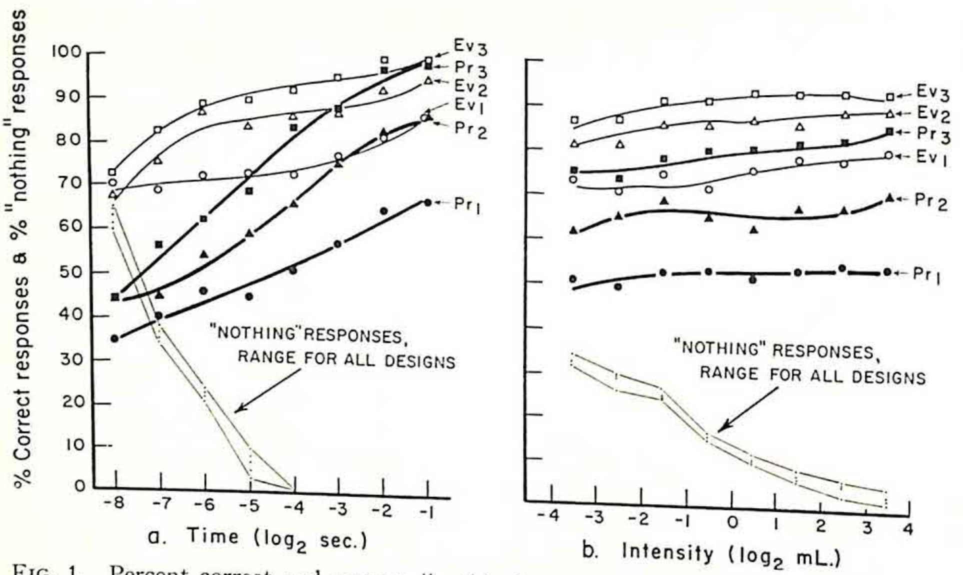


FIG. 1. Percent correct and percent "nothing" responses for each of the six patterns as a function of exposure time (a) and as a function of intensity (b). (Curves fitted with cubic equations.)

subtracting the product of 560 and the percent NO responses from 560 and treating the resulting number as the denominator for computing percent correct responses. The curves shown in Fig. 1 (a and b) were obtained by fitting cubic equations to the data, using the method of least squares.

Figure 1a shows the percent correct responses for each design as a function of ET averaged over all I's and *S*s. As expected, the percent correct responses at each ET decreased from Design Pr3 to Pr2 to Pr1, and accuracy to Pr designs increased as a function of ET.

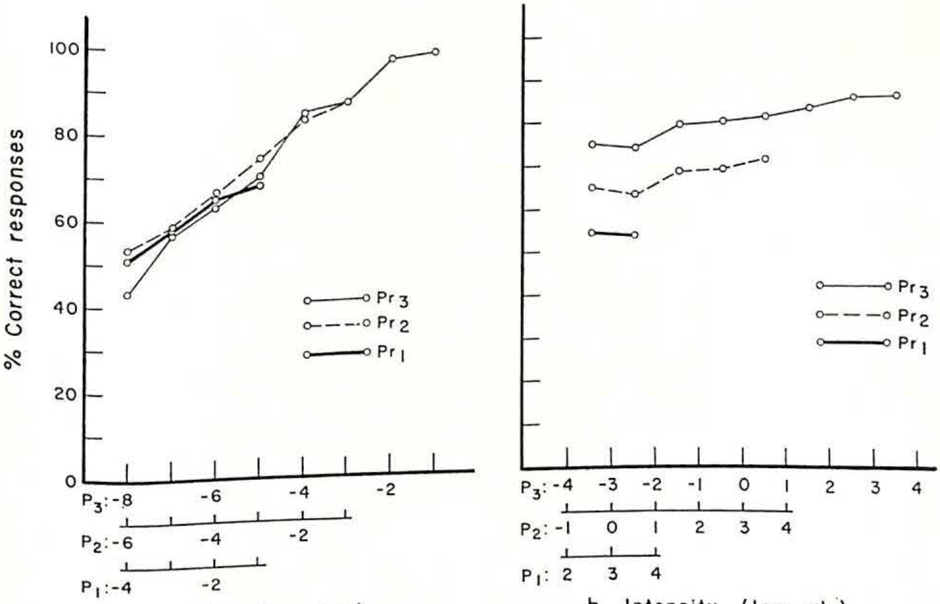
Figure 2a examines the functional equivalence of the effect of RSD and ET on accuracy which was expected from the assumption of the interaction of spatial differentiation and ET. To facilitate this comparison, the Pr designs were constructed so that RSD increases from Pr1 to Pr2 to Pr3 on the same log basis as do I and ET. Thus, as can be seen from Table 1, the difference obtained by subtracting the distance between pairs from that within pairs is 2 mm., 4 mm., and 8 mm. for Designs Pr1, Pr2, and Pr3, respectively. Accordingly, these differences, which are the values of RSD for these designs, increase on a $\log_2$ scale. The considerable superimposition of curves in Fig. 2a was obtained by transposing the Pr design curves from Fig. 1a, but taking the initial data point at .004 sec. ($\log_2 = -8$), .016 sec. ($\log_2 = -6$), and .064 sec. ($\log_2 = -4$) for Pr3, Pr2, and Pr1, respectively. Thus, RSD and ET do not show a one-to-one relationship, but the results indicate that when ET is quadrupled (e.g., .004 sec. to .016 sec.) for each successive halving of RSD (e.g., Pr3 vs. Pr2) a RSD-ET equivalence is approached in terms of percent correctly perceived.

Figure 1a also shows the overall effect of ET on Ev designs and permits a comparison of results for each set of Ev-Pr designs (i.e., Ev1 vs. Pr1; Ev2 vs. Pr2; Ev3 vs. Pr3). As expected, accuracy to the Ev design of a given set was greater than accuracy to the Pr design in that set, especially at short ETs. Contrary to expectations, however, the curves for Ev designs tend to follow the order of accuracy for Pr designs and show a definite growth function, especially at short ETs. Since discrimination of Ev designs was presumed to temporally precede discrimination of Pr designs, there was no reason to expect that Ev designs should ever be seen as paired. Occasional errors would be ascribed to confusion, such as accidentally hitting a wrong key, which should be independent of ET or I so that Ev design accuracy curves were expected to be horizontal and superimposed. In spite of this deviation from expectations, the maximum differences in accuracy between Ev and Pr designs of each set still occurred at brief ETs as postulated by the differentiation hypothesis.

RSD-I.—Figure 1b gives percent correct responses for each design as a function of I averaged over all ETs and *S*s. It is clear from the horizontal curves that changes in I had little effect on accuracy so that the ordinal relations between designs reflect primarily the average effect of ET shown in Fig. 1a.

It is obvious that the relationship found for ET and RSD as shown in Fig. 2a cannot be obtained for I and RSD since the I curves shown in Fig. 1b are parallel and nonoverlapping and, therefore, cannot be superimposed by an I scale shift.

It was found, however, that by transposing the Pr curves in Fig. 1b and taking initial data points at .093 mL. $(\log_2 = -3.5)$, .74 mL. $(\log_2 = -.5)$, and 5.92 mL. $(\log_2 = 2.5)$ for Pr3, Pr2, and Pr1, respectively, a constant difference of about 5% was present between all remaining data points of the three designs as shown in Fig. 2b. This finding suggests that, while I, unlike ET, cannot compensate for decreases in spatial differentiation,



a. Time $(\log_2 \text{sec.})$ b. Intensity $(\log_2 \text{mL.})$

FIG. 2. Percent correct responses for paired designs as a function of exposure time (a) and intensity (b). (Distributions are plotted at values of time and intensity which give the closest approximation to D × ET = k and D × I = k, respectively.)

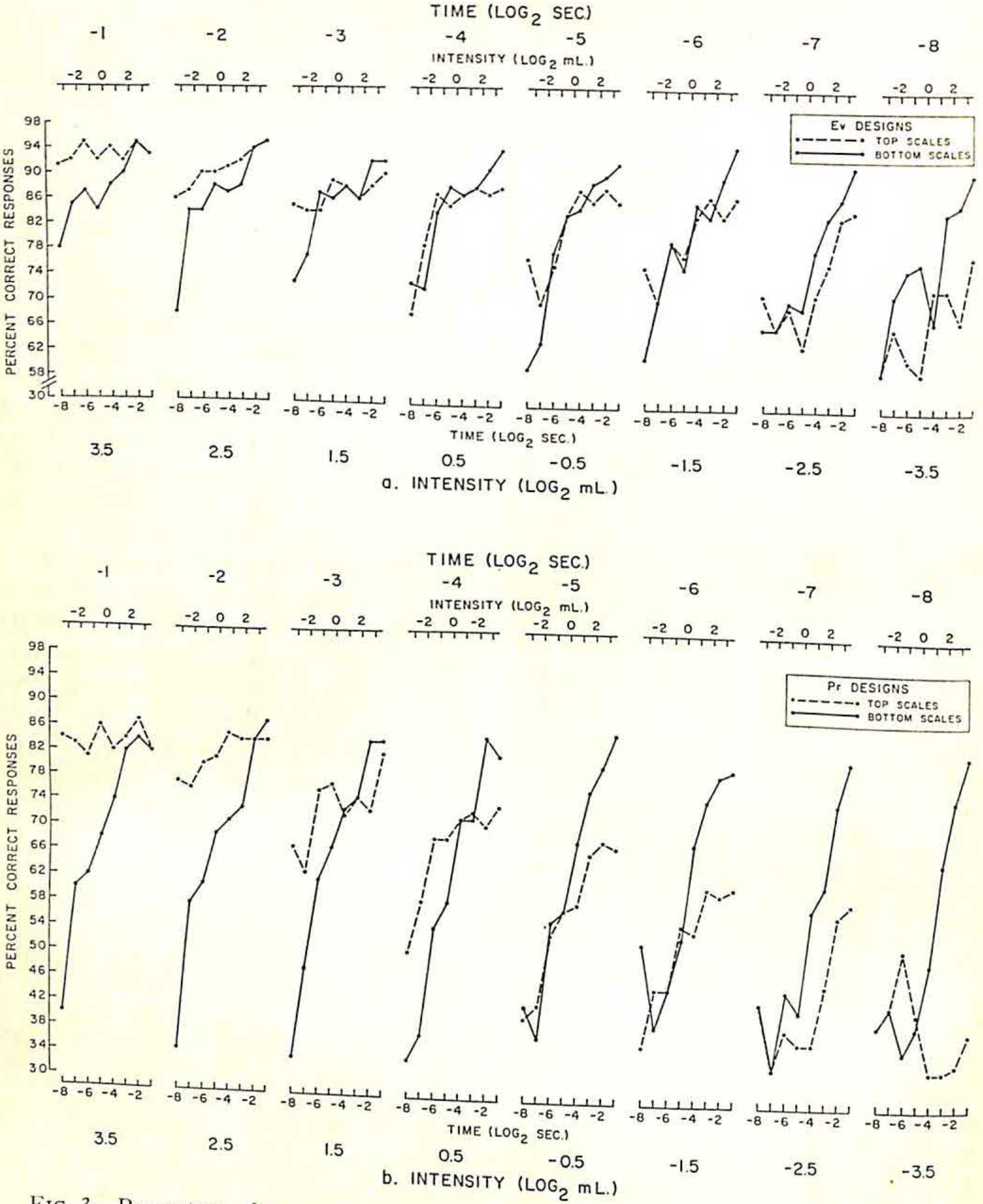


FIG. 3. Percentage of correct responses as a function of exposure time within each intensity (solid lines) and as a function of intensity within each exposure time (dashed lines). (Each distribution is plotted to permit comparison of complementary I-ET log units, so that if I × T = k, solid and dashed lines should be superimposed. Data for Ev designs are shown in a, data for Pr designs are in b.)

I does have a slight effect on discriminability in a regular relationship to RSD.

I-ET relations.—Figure 3 (a and b) shows I-ET interactions most clearly. Percent correct responses were again based on the ratio of incorrect to the sum of correct and incorrect responses, excluding NO responses. There was a total of approximately 210 exposures at each point of these two graphs. The number of NO responses can be seen in Fig. 4 and can be used to compute the actual *N* at each point of Fig. 3 (a and b). Standard deviations at individual points ranged in percent from 1.45 at 11.84 mL. and 8 msec. to 8.51 at .09 mL. and 8 msec. for Ev designs and from 2.35 at 5.43 mL. and 8 msec. to 7.84 at .18 mL. and 4 msec. for Pr designs. The median *SD* for Ev designs was 2.37%, and for Pr designs it was 3.36%. Standard deviations at each point were estimated by $\sqrt{PQ/N}$, where P = percent correct responses, Q = percent incorrect responses, and N = sum of correct and incorrect responses.

The bottom scales of these figures are constructed to show the effect of ET within each I, going from the brightest I at the left to the dimmest I at the extreme right with ET increasing from left to right within each I block. Results are shown with solid lines. It is clear that the increase in accuracy as a function of ET at each I block is about the same. The scales on the top of Fig. 3 (a and b) were constructed to show the effect of I within each ET, with the longest ET at the extreme left, and the shortest ET at the extreme right. Intensities increase from left to right within each ET block, and results are plotted with dashed lines. Both figures, but especially Fig. 3b, show that ET is the limiting factor for accuracy, since given long ETs (extreme left) changes in I had no effect, and at the briefest ET (extreme right) accuracy was not improved by increasing I. At intermediate block values, the effect of ET was always greater than that of I, especially for Pr designs (Fig. 3b). These results held for all designs, although the average magnitude of accuracy shifted as a function of RSD (see Fig. 1a).

Figure 3 (a and b) was also constructed to assess possible I-ET reciprocity, since points on dashed and solid lines represent equal I-ET products (i.e., $\log I + \log t = k$). Thus, for example, the first point at the extreme left of each graph in the solidly drawn curves represents the accuracy of discrimination at the brightest I and the fastest ET. The complementary point on the curve drawn with dashes reflects the accuracy of discrimination at the longest ET and dimmest I. By the nature of this method of graph construction, two points in each block are always superimposed, since they are the same I-ET combinations. It is only when both ET and I are intermediate values that anything approaching I-ET reciprocity can be discerned. This tendency toward reciprocity appears somewhat greater for Ev (Fig. 3a) than for Pr designs (Fig. 3b).

In addition to greater I-ET reciprocity, accuracy for Ev designs also changed over a narrower range than that for Pr designs. Further, while indications of I-ET reciprocity are absent at extreme blocks, some effect of I on accuracy is present even at the fastest ET for Ev but not for Pr designs.

In order to help evaluate the effect of I and ET on discrimination independent of spatial distance variables, a supplementary experiment was carried out with the same *S*s on completion of the main study. They were shown either a centrally located 10-mm. black dot or a blank field, and required to indicate what was shown by depressing one of two keys corresponding to the stimuli. Each of the 5 *S*s was given two exposures of each stimulus at each of the I-ET values used in the main study, so that there was a total of 10 exposures of each I-ET combination for each stimulus

TABLE 2

Errors to Blank (B) and Dot (D) Stimuli as a Function of I and ET

ET (Sec.)	I (ML.)					
	.09		.19		.38	
	D	B	D	B	D	B
.004	9	0	8	1	7	0
.008	7	1	5	1	1	0
.016	1	1	0	0	0	0

Note.—Total errors to all other ET-I combinations not shown were 2 (D) and 8 (B).

for all *S*s combined. The results are shown in Table 2. It is clear that errors to the dot predominate and are concentrated at dim and fast exposures.

NO responses, S, *and practice effects.* Figure 1 (a and b) indicates that the frequency distributions of NO responses are about the same for all designs. An analysis of variance comparing total frequency of NO responses for each design yielded an $F(5, 20) = 2.31$, $p > .05$. Figure 4 shows the frequency of NO responses for each ET as a function of I. Since the NO responses occurred with almost exactly the same frequency for all designs only the combined curves for Pr designs are shown here. Each point is based on 210 responses. It is clear that NO responses were concentrated at short and dim exposures. Since the single dot in the supplementary study was also rarely seen at these exposures, it is probable that NO responses can generally be accepted at face value. No ETs above .064 sec. are shown since there was only a total of three NO responses above that point. It is interesting to note that when successively longer ET curves in Fig. 4 are shifted back about two log units along the I scale, for each log unit of ET the curves are

TABLE 3

Number of Correct and Incorrect (*N*) and Percent Correct Responses as a Function of *S*, Intensity, and Exposure Time for Ev Designs

Intensity	Subject	Exposure Time											
		Fast ET (.004–.032 Sec.)						Slow ET (.064–.512 Sec.)					
		Ev1		Ev2		Ev3		Ev1		Ev2		Ev3	
		N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Bright (1.94–11.84 mL.)	AP	164	87	173	90	217	77	254	71	211	87	213	99
	CW	238	73	217	84	226	84	240	94	221	95	223	99
	LP	176	79	183	89	180	94	209	82	229	94	254	100
	JH	199	61	168	71	206	75	239	85	195	92	253	92
	GF	196	76	183	89	191	90	228	83	205	90	225	98
Quadrant mean		($N = 195$; % = 81.6)						($N = 227$; % = 91.0)					
Dim (.09–.74 mL.)	AP	90	79	83	87	66	80	214	77	221	88	225	96
	CW	213	62	197	62	200	71	240	85	222	95	225	98
	LP	88	77	89	84	75	83	205	76	228	92	257	95
	JH	93	43	59	61	77	69	230	76	194	86	243	91
	GF	117	80	104	85	125	88	213	68	210	87	223	95
Quadrant mean		($N = 112$; % = 75.6)						($N = 223$; % = 86.8)					

TABLE 4

NUMBER OF CORRECT AND INCORRECT (N) AND PERCENT CORRECT RESPONSES AS A FUNCTION OF S, I, AND ET FOR Pr DESIGNS

Intensity	Subject	Exposure Time											
		Fast ET (.004–.032 Sec.)						Slow ET (.064–.512 Sec.)					
		Pr1		Pr2		Pr3		Pr1		Pr2		Pr3	
		N	%	N	%	N	%	N	%	N	%	N	%
Bright (1.48–11.84 mL.)	AP	176	24	174	31	170	70	223	58	220	79	224	98
	CW	234	58	215	79	220	85	259	82	226	97	226	99
	LP	177	40	184	55	214	78	207	63	219	84	249	97
	JH	198	59	132	71	202	83	242	57	189	68	256	89
	GF	204	33	187	24	189	30	224	50	211	72	227	85
Quadrant mean		(N = 192; % = 54.8)						(N = 227; % = 76.6)					
Dim (.09–.74 mL.)	AP	91	37	110	32	69	62	224	49	231	73	226	92
	CW	190	53	195	64	198	66	237	73	221	95	225	97
	LP	81	30	90	53	81	64	211	58	228	78	261	87
	JH	74	51	59	63	66	85	239	57	192	69	256	96
	GF	124	23	103	19	120	13	224	49	206	55	219	68
Quadrant mean		(N = 107; % = 47.4)						(N = 227; % = 73.2)					

superimposed within a range of about 10% in accuracy. This indicates that when accuracy of perception is disregarded some I-ET reciprocity is obtained.

Tables 3 and 4 combine the eight I and eight ET values into brighter and dimmer and faster and slower halves, respectively, and give the total number of correct plus incorrect responses (N) and the percent correct responses for each S for each design at each combination of I and ET. While the principal purpose of these tables was to present individual differences, they also provide a more convenient but less detailed comparison than graphic presentations of N and percent correct responses to changes in the three stimulus parameters. As can be seen most clearly from the overall quadrant means, the role of I is much more important for N than for percent correct responses. While the accuracy and number of responses varied considerably from S to S, the effects of RSD, I, and ET were usually in the same direction for each S as that shown in the combined data (Fig. 1, a and b).

In order to assess practice effects, each S's 40 sessions were divided into quartiles, according to the chrono-

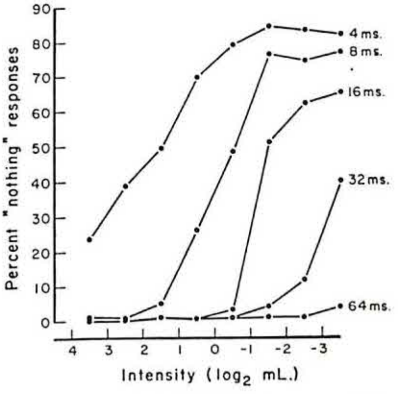


FIG. 4. Percent "nothing" responses as a function of exposure time and intensity.

logical order of sessions to see if an increase in percent correct responses could be found from the first through the fourth quartile. The obtained $F(3, 12) = .11$ ($p > .05$) indicated no change in accuracy with practice. In fact, percent correct in the first and last quartile was about the same. This indicates that the initial practice sessions provided sufficient warm up for this task.

Discussion

The results of this study provide clear evidence that, for the stimuli used here, perception occurs largely as a result of spatial differentiation over time. This was shown not only by the increase in accuracy to Pr designs as a function of ET, but also by the increase in accuracy as a function of increasing RSD (Pr1, Pr2, Pr3), so that ET and RSD tended to have functionally equivalent effects on accuracy (Fig. 2a). Emphasis on RSD is important, since increases only in the absolute distance between dots (Ev1 < Ev2 < Ev3), resulted in accuracy curves which tended to be parallel to each other and to the abscissa, so that, unlike the Pr design curves, the Ev design curves could not be superimposed.

The finding of a much greater effect of ET than I on accuracy of discrimination in these stimuli (Fig. 3, Tables 3 and 4), is in accord with findings of Leibowitz and Bourne (1956) for the effect of I and ET on shape constancy. When shape judgments were plotted as a function of increasing I with the fastest ET (.01 sec.) as the parameter, the resulting curve was quite flat and very close to the law of retinal image. When shape judgments were plotted as a function of increasing ET with the dimmest luminance (.01 mL.) as the parameter, there was a regular, seemingly linear, increase toward shape constancy. The authors note that constancy tended to be destroyed by decreases in stimulation because constancies presumably depend on the presence in the visual field of stimuli in addition to the discriminative stimulus which cannot be discriminated at low energy levels. While the nature of these "additional" stimuli was not specified, it is clear that ET played a more decisive role than I in making such stimuli effective for shape constancy.

While the stimuli used in this study cannot be considered directly analogous to those involved in shape constancy, it seems reasonable to consider RSD as indicating characteristics of the designs "additional" to the dots for the perception of groupings, in the same sense in which size and distance may be additional cues for size constancy. The essential point is that the relation among stimuli is critical for the response in both cases.

The NO responses or the converse—the sum of accurate and inaccurate responses—are analogous to responses in a detection task, where only the identification of the presence or absence of a stimulus is considered. Under these conditions I was found to be a more critical stimulus variable than it was in response accuracy. Further, since there were about the same number of NO responses for all designs, detection was shown to be independent of spatial relations. These findings are in line with the results of other studies showing I-ET reciprocity for such detection tasks as dot numerosity (Hunter & Sigler, 1940), grating acuity (Graham & Cook, 1937), absolute threshold (Graham & Magaria, 1935), and "frequency of seeing" studies (Leibowitz, Myers, & Grant, 1955a, 1955b). While it might appear that dot numerosity identification might require more complex discrimination than merely identifying the presence of a stimulus, Schlosberg's (1948) formulation indicates that it is possible to predict the Hunter and Sigler results fairly well by considering the contribution of each dot to the discrimination as an independent event, reducing the problem essentially to a set of detection tasks. The Leibowitz et al. (1955a) study supports this formulation for small numbers of dots, as does a recent study by Restle, Rae, and Kiesler (1961).

These considerations suggest that a two-phase temporal process may occur in the perception of complex displays. The first phase presumably involves only the detection of the presence or absence of stimuli and depends on the total energy of stimulation (I × ET). When the task requires that different stimuli be related to each other, or where, as in this study, the spatial relations among elements within a stimulus array are relevant, the ET required for detection is insufficient for accurate identification. Once the presence of stimuli has been detected, a second phase, discrimination, follows. In this phase, the amount of *additional* ET required for accurate discrimination appears to depend on the amount of spatial differentiation in the design (see Fig. 2a).

Acceptance of this formulation requires a revision of the usual conception of I and ET as reciprocal up to some critical ET (e.g., Boynton, 1961, p. 740). The present findings and the pertinent literature noted above suggest that reciprocity fails to hold even at ETs in the millisecond range when perceptual accuracy depends on the relational characteristics of stimuli. Also, it is important to note that no postulation of cognition such as counting or thinking, need be presumed in the temporal development of discrimination for these orders of magnitude since enough information is obtained within 64 msec. to result in better than 70% accuracy for P designs at all but the two dimmest intensities (see Fig. 3b). Even higher accuracy was obtained for Ev designs within this ET limit (Fig. 3a). Since cortical transmission times are probably much slower than 64 msec. (Lansing, Schwartz, & Lindsley, 1959), cognitive evaluation of the stimuli may occur after the *cessation* of the exposure and prior to the response, but this is not likely to play a role in the ET-RSD relationship, which largely determines the information for such evaluation.

Considering I-ET relations to be more critical for detection than for accurate discrimination of more complex displays suggests a reason why greater I-ET reciprocity was found for Ev than Pr designs (Fig. 3). Since no differential spatial cues need to be taken into account for the identification of these designs, Ev design identification is probably closer to, and more likely to overlap with the detection phase of perception, as is also indicated by the briefer ETs necessary for identification of Ev than of Pr designs.

REFERENCES

BENDER, L. A visual motor gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. res. Monogr.*, 1938, No. 3.

BOYNTON, R. M. Some temporal factors in vision. In W. A. Rosenblith (Ed.), *Sensory communication.* New York: Wiley, 1961. Pp. 739–755.

FLAVELL, J. A., & DRAGNUS, J. A. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54,** 197–217.

GRAHAM, C. H., & COOK, C. Visual acuity as a function of intensity and exposure time. *Amer. J. Psychol.*, 1937, **49,** 654–661.

GRAHAM, C. H., & MARGARIA, R. Area and the intensity-time relation in the peripheral retina. *Amer. J. Physiol.*, 1935, **113,** 299–305.

HUNTER, W. S., & SIGLER, M. The span of visual discrimination as a function of time and intensity of stimulation. *J. exp. Psychol.*, 1940, **26,** 160–179.

KASWAN, J. W. Tachistoscope exposure time and spatial proximity in the organization of visual perception. *Brit. J. Psychol.*, 1958, **49,** 131–138.

LANSING, R. W., SCHWARTZ, E., & LINDSLEY, D. B. Reaction time and EEG activation under alerted and nonalerted conditions. *J. exp. Psychol.*, 1959, **58,** 1–7.

LEIBOWITZ, H., & BOURNE, L. E., JR. Time and intensity as determiners of perceived shape. *J. exp. Psychol.*, 1956, **51,** 277–281.

LEIBOWITZ, H. W., MYERS, N. A., & GRANT, D. A. Frequency of seeing and radial localization of single and multiple visual stimuli. *J. exp. Psychol.*, 1955, **50,** 369–374. (a)

LEIBOWITZ, H. W., MYERS, N. A., & GRANT, D. A. Radial localization of a single stimulus as a function of luminance and duration of exposure. *J. Opt. Soc. Amer.*, 1955, **45,** 76–78 (b)

RESTLE, F., RAE, J., & KIESLER, C. The probability of detecting small numbers of dots. *J. exp. Psychol.*, 1961, **61,** 218–222.

SCHLOSBERG, H. A probability formulation of the Hunter-Sigler effect. *J. exp. Psychol.*, 1948, **38,** 155–167.

SMITH, G. J. W., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59,** 167–176.

(Received March 26, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 124–130

PERCEPTUAL RESPONSES AS A FUNCTION OF THE SEQUENTIAL PROPERTIES OF MULTIPLE VISUAL STIMULI[1]

LEE W. GREGG AND HARRY W. KARN

Carnegie Institue of Technology

Circles located at the points of a triangle were differentially loaded with runs of dots and no dots averaging 2, 4, and 8 in length. Different combinations of loading and location were tachistoscopically presented to 3 experimental groups with a control group receiving random dot and no-dot combinations at all locations. Experimental groups performed significantly better in terms of error reduction than did the control group. However, variance analysis failed to show an effect of run length, per se. Despite the statistical results, an effect of run length is suggested by an analysis in which straight lines were fitted to the data by the method of least squares.

Under appropriate conditions of presentation, the more frequent of two sequentially presented events can be discriminated even though the rate at which the events occur precludes tallying after each occurrence. In addition to showing that the accuracy of judgment increases as the frequency differential between the two events increases, Erlick (1961) found that there was a slight overestimation for the event having a higher degree of clustering or repetitiveness.

In a more complex perceptual task where the states of several simultaneously presented stimuli are reported following each tachistoscopic presentation of these stimuli, sequential properties of the separate elements might provide learnable cues that could modify performance of the task. Karn and Gregg (1961a, 1961b) reported that changes in perceptual responses do occur when the values of two of three stimulus elements are invariant throughout the series of presentations. The stimuli were separated spatially so that it was impossible to bring all elements into foveal regard at the viewing distance and duration of exposure that were used. Although the nearly perfect performance in reporting the state of the stimulus that one might expect in a situation like this did not come about, perceptual learning attributable to the stable "loading" was clearly demonstrated. Whether a more transilient loading will produce learning effects is the main question posed for the present study.

The sequential characteristics of the stimulus elements are defined by a variable, Average Length of Run. This variable, like the clustering in the Erlick (1961) experiment determines with what probability a given stimulus value will be reported on the next presentation even though the frequencies of the binary values are the same for the total series.

METHOD

Subjects.—The *S*s were 40 college students, 10 in each of three experimental groups, and 10 in a control group.

Apparatus and procedure.—Three circles located at the points of an imaginary equilateral triangle were simultaneously presented to *S* on a 4 × 5 ft. screen. Each side of the triangle was 30 in. long subtending at *S*'s

---

[1] This research was supported by the Air Force Office of Scientific Research of the Office of Aerospace Research under Contract No. AF 49(638)-770.

viewing distance approximately 15° of visual angle. The outside diameter of each circle was 1 in., corresponding to a visual angle of 2′ of arc. The discrimintation required of *S* was the presence or absence of a dot in a circle; the dot subtending about 1′ of arc.

Three different stimulus configurations (Cs) were used with the experimental *S*s. In C_1, a long run of circles with dots present followed by a long run with dots absent appeared at the apex of the triangle (Location A); the average length of these runs was 8. In the circle at the end of the base of the triangle to *S*'s right (Location B) dot and no-dot runs averaged 4 in length. Short runs of 2 dot and 2 no-dot circles appeared at the other end of the triangle's base to *S*'s left (Location C). In C_2 medium length runs appeared at the apex, short runs to *S*'s right, and long runs to *S*'s left. In C_3 short runs were at the apex, long runs to the right, and short runs to the left. For the control group, dot and no dots appeared at random in all locations.

For runs of Average Length 8, the probability that a given value of a stimulus will follow that same value on the next presentation is .4375; the probability that the value of the stimulus will change is .0625; the difference in these dependent probabilities is .3750. For runs of Length 4, the dependent probability difference is .2500. For runs of Length 2, where each of the four separate dependent probabilities is .2500, the difference is 0; hence for the shortest runs the occurrence of a dot or no-dot stimulus is equally likely on any presentation. The random sequences in all stimulus locations for the control group similarly yield a probability difference of 0.

Each group of experimental and control *S*s received 128 presentations of the stimulus complex at an exposure time of .1 sec. per presentation. A response panel with three toggle switches was attached to an armchair in which *S* was seated. Each switch had a neutral position and could be thrown forward to indicate the presence of a dot in a circle and back toward *S* to indicate the absence of a dot.

The experimental room was dimly illuminated at all times by a single lamp located above and to the rear of *S*. The screen brightness was approximately 5 mL.

The instructions to *S* were:

> On this screen I am going to flash three circles located at the points of an imaginary triangle. Sometimes one or more of the circles will contain a small dot in the center. Your job is to let me know which circles contain dots and which are blank. You will indicate the presence or absence of a dot in each circle by using the switches on the armchair. For the circle to your right, use the switch to your right; for the circle at the apex use the center switch; for the circle to your left use the switch to your left. If the circle contains a dot, throw the switch forward. If there is no dot, pull the switch toward you. The duration of the flashes may be so short that you cannot be sure of the presence or absence of a dot in all of the circles. If you are not sure, guess. Give a response to each circle every time. I shall operate the apparatus in the rear of the room. I'll give you a ready signal, then flash the circles after which you are to make three switch throwings—one for each circle. Throw the appropriate switch forward if there is a dot, backward if there is no dot. Do the best you can.

RESULTS

Error scores were computed for the separate blocks of 64 presentations for each of the three stimulus locations and two types of error. Table 1 presents the mean errors per block for the four groups of 10 *S*s each and the mean total errors. In the table, P errors are errors of omission for which *S*s report the dot absent when it is in fact present; N errors are errors of commission—reports that the dot is present when it is not.

An overall analysis of variance of the total errors was performed. The main effect of Blocks was significant, $F(1, 36) = 14.42$. The main effect of Groups was not significant; however, the B × G interaction did reach significance at the .01 level, $F(1, 36) = 4.41$. The significant interaction term indicates that the magnitude of the differences in error reduction from Block 1 to Block 2 vary for the four groups. These mean differences were 19.1, 23.1, and 21.6 errors for the experimental groups C_1, C_2, and C_3, respectively. Mean error reduction for the control group, however, was only 8.4 errors. It is apparent that significant interaction comes about

TABLE 1

Mean Errors per Block of 64 Presentations

Group	N	Block	Locations						Mean Total
			A		B		C		
			P	N	P	N	P	N	
C_1 (8-4-2)	10	1	5.4	13.9	8.7	14.8	1.5	13.1	57.4
		2	4.6	8.1	6.6	12.0	1.8	5.2	38.3
		Mean	5.0	11.0	7.2	13.4	1.6	9.2	
C_2 (4-2-8)	10	1	7.8	16.3	10.9	16.0	1.8	15.0	67.8
		2	4.3	11.9	7.9	13.8	.8	6.0	44.7
		Mean	5.0	14.1	9.4	14.9	1.3	10.5	
C_3 (2-8-4)	10	1	10.6	14.7	12.6	11.3	1.4	13.4	64.0
		2	5.2	11.7	9.5	7.6	.6	7.8	42.4
		Mean	7.9	8.2	11.0	9.4	1.0	10.6	
Control	10	1	8.3	12.8	11.8	9.3	3.2	11.9	57.3
		2	5.5	11.7	10.9	9.4	2.5	8.9	48.9
		Mean	6.9	12.2	11.4	9.4	2.8	10.4	

because the magnitude of the error reduction for the experimental groups is approximately 2.5 times that for the control group.

A separate analysis of variance of the errors for each stimulus location and type of error for the control group data is presented in Table 2. The main effects of Blocks and Locations are significant as are the L × E and B × L × E interactions. The significant effect of Blocks indicates that the reduction in mean errors from 9.4 to 8.2 errors, although small is reliable. The significant effect of Location is reflected in the error scores of 6.6 at Location C; 9.6 at Location A; and 10.4 at Location B. The L × E interaction reflects the small number of P errors at the preferred location.

TABLE 2

Analysis of Variance for the Control Group

Source	*df*	*MS*	*F*
Blocks (B)	1	58.80	7.50*
Location (L)	2	154.53	4.81*
Error type (E)	1	396.03	3.03
B × L	2	7.53	.51
B × E	1	.14	.00
L × E	2	250.11	5.51*
B × L × E	2	11.41	4.35*
Ss	9	85.03	
Ss × B	9	7.84	
Ss × L	18	32.11	
Ss × E	9	130.85	
Ss × B × L	18	14.70	
Ss × B × E	9	38.98	
Ss × L × E	18	45.40	
Ss × B × L × E	18	2.62	
Total	119		

* $p < .05$.

The analysis of variance presented in Table 3 is for the three experimental groups. By treating these groups separately from the control group, the analysis can take the form of a Latin square, and thus permit the assessment of the variable Run Length which in each configuration has one of the three possible values of the variable associated with one of the three locations. Locations, Blocks, and Error Type are significant main effects. The B × E and B × E × L interactions are also significant. The main effect of Run Length is not significant nor are any of the interactions involving this source. For the separate locations, the errors follow the same pattern as the control group: 5.7 errors at Location C, 9.3 errors at Location A, and 11.0 errors at Location B. There are significantly fewer P errors, 5.6, than N errors, 11.7.

DISCUSSION

Perceptual learning involving improvement in reporting the state of binary-valued stimuli is enhanced for stimulus complexes where the sequential properties of the stimuli vary from stimulus location to stimulus location. The control group (which received presentations of a random, short-sequence average length of 1.88 at each of the separate locations) showed significant improvement in perceptual performance but not to the extent exhibited by the experimental groups. As in the earlier studies (Karn & Gregg, 1961a, 1961b), significant location and error-type differences were observed. Fewer errors were made at Location C than at the other two locations; P errors were less than N errors, reflecting a predominant response bias to report the presence of a dot in the circles.

The Latin square analysis for the experimental groups does not provide direct evidence for the expectation that average length of run would provide a learnable cue if the runs were long (8) but would not affect performance if short (2). Nevertheless, some aspect of the complex stimulus display must have made it possible for the *S*s in the experimental groups to reduce their errors to a greater degree than in the control.

One way that performance in the perceptual task could improve is to identify the longer run location, note the value of the stimulus, and, at least for some number of subsequent presentations, report that same value while attending to the other stimulus elements that fluctuate more rapidly. If *S* did in fact learn something of the sequential properties associated with the long runs of Length 8, then accuracy of report for the stimulus elements of runs of Length 2 or 4 could increase. It is conceivable that no gross statistical analysis over large blocks of trials could detect differential effects of run length for configurations of the sort used in the present experiment. This kind of perceptual strategy was postulated in the earlier studies (Karn & Gregg, 1961a, 1961b) where the stimulus values were completely stable for two of the three elements.

It is apparent that a more detailed analysis of *S*s' momentary responses is required. In Fig. 1, ordered blocks of runs of Length 8 were used as the base for following the sequence of *S*s' responses: The possibility that differential loading in the form of length of run contributes to the results of the present experiment is suggested by this analysis. For each of the three experimental groups, straight lines were fitted by the

TABLE 3

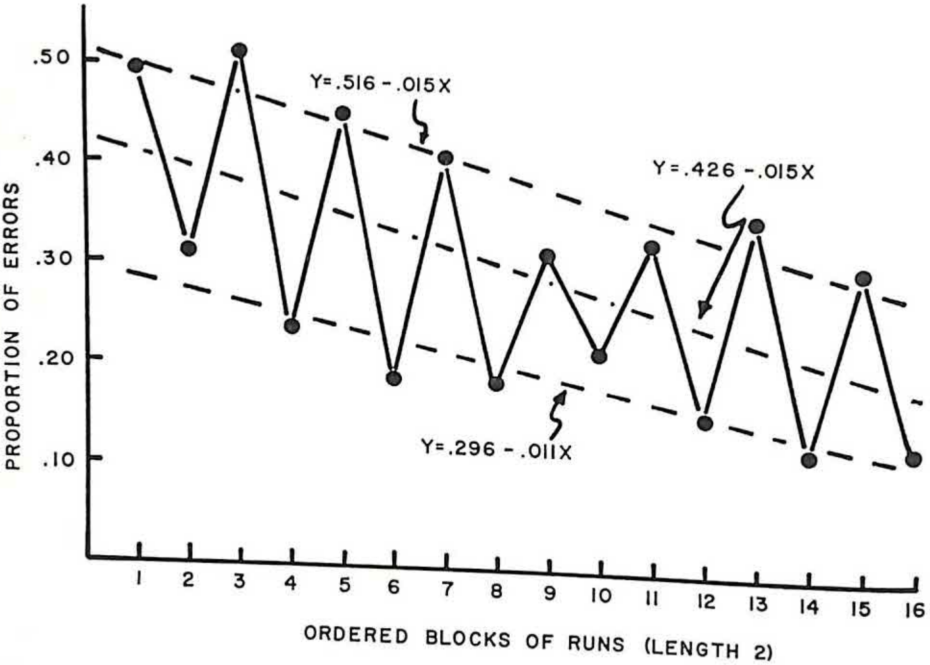
ANALYSIS OF VARIANCE FOR EXPERIMENTAL GROUPS

Source	*df*	*MS*	*F*
Location (L)	2	892.77	24.79*
Configuration (C)	2	60.27	
Run length (R)	2	52.34	
Square uniqueness (SU)	2	5.92	
*S*s within Cs	27	72.05	2.00
Latin square error	216	36.01	
Blocks (B)	1	1130.68	31.39*
B × L	2	12.17	
B × C	2	3.40	
B × R	2	3.68	
B × SU	2	6.09	
B × *S*s	27	13.94	
Error type (E)	1	3385.60	29.84*
E × L	2	236.06	
E × C	2	77.81	
E × R	2	81.70	
E × SU	2	52.35	
E × *S*s	27	113.44	
B × E	1	185.89	5.16*
B × E × L	2	96.11	3.58*
B × E × C	2	23.06	
B × E × R	2	16.12	
B × E × SU	2	4.28	
B × E × *S*s	27	26.82	
Total	359		

* $p < .05$.

least squares method to the P errors, N errors, and combined data points. An inspection of the graphs and the slope and intercept constants presented in Table 4 indicates an effect of Run Length on performance during the perceptual task. Only for runs of Length 8 do the lines forming an envelope converge. For the shorter runs, 4 and 2, the envelopes are approximately uniform; and for the control group, the envelope appears to diverge slightly. The overall slope for the three experimental groups, .014, is about twice that for the control group.

No statistical theory of visual de-



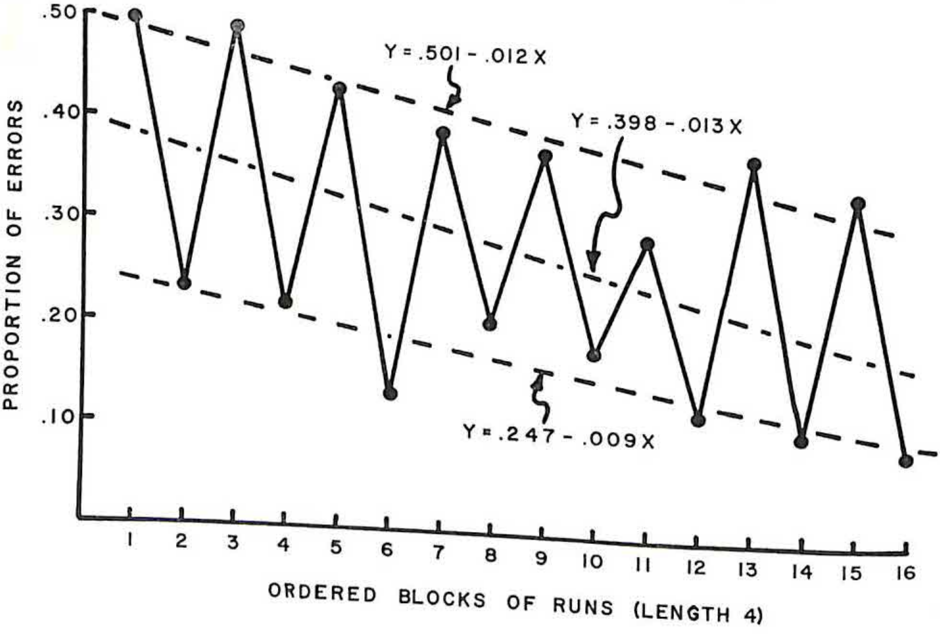
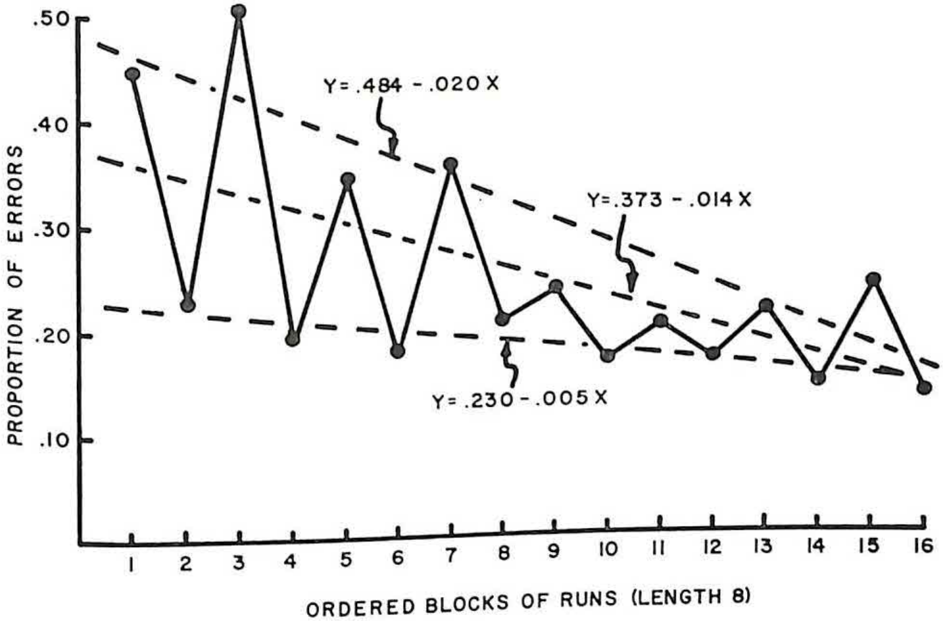


FIG. 1.—Continued



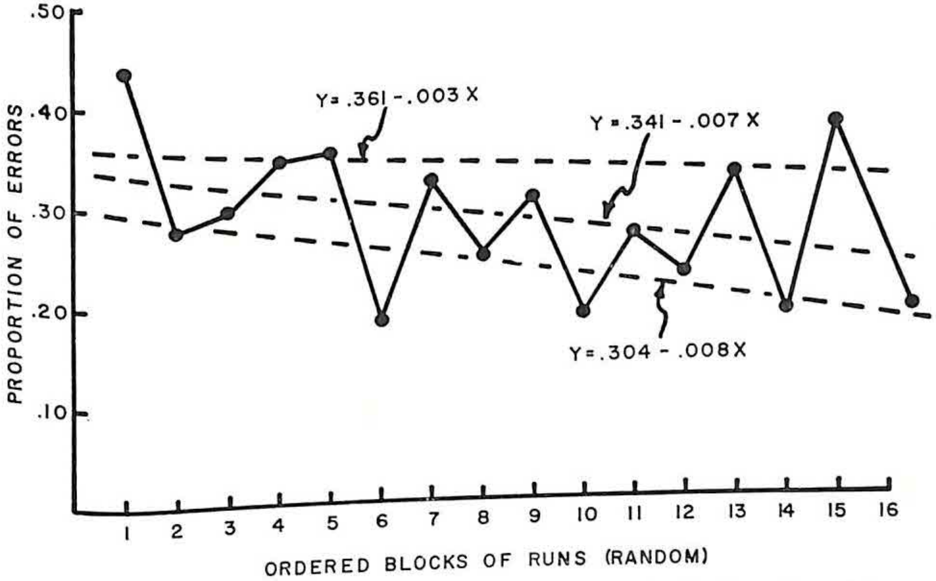


FIG. 1. Relationship between mean proportion of errors and ordinal position of alternating blocks of P and N runs. (Even-numbered blocks are for P errors; odd-numbered blocks are for N errors. Only in the case of runs of Length 8 are the blocks sequential. Broken lines are least squares, linear fits for the P errors, N errors, and combined data.)

tection is elaborate enough to predict quantitatively responses to simultaneously presented stimuli, even though it is reasonably certain that response biases of the sort implied by Tanner and Swets (1954; Swets, 1961) must be present What is required of theory is a more precise specification of the mechanism of fixation, attention, and report tha act together to generate complex be

TABLE 4

LEAST SQUARE SOLUTIONS

Run Length	Errors	Intercept	Slope
2	P	.296	.011
	N	.516	.015
	Total	.426	.015
4	P	.247	.009
	N	.501	.012
	Total	.398	.013
8	P	.230	.005
	N	.484	.020
	Total	.373	.014
Control	P	.304	.008
	N	.361	.003
	Total	.341	.007

havioral consequences. Until recently, such determinism was not feasible. With the advent of computer simulation techniques (Laughery & Gregg, 1962; Newell, Shaw, & Simon, 1958; Reitman, 1959) however, it is reasonable to suppose that progress in relating multiple determinants of perceptual responses will be made.

REFERENCES

ERLICK, D. E. Judgments of the relative frequency of a sequential series of two events. *J. exp. Psychol.*, 1961, **62**, 105–112.

KARN, H. W., & GREGG, L. W. Acquisition of perceptual responses as a function of loading, location, and repetition. *J. exp. Psychol.*, 1961, **62**, 62–69. (a)

KARN, H. W., & GREGG, L. W. Supplementary report: Effects of instructions on the perception of multiple targets. *J. exp. Psychol.*, 1961, **62**, 533–534. (b)

LAUGHERY, K. R., & GREGG, L. W. Simulation of human problem-solving behavior. *Psychometrika*, 1962, **27**, 265–282.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Elements of a theory of human problem solving. *Psychol. Rev.*, 1958, **65**, 151–166.

REITMAN, H. R. Heuristic programs, computer simulation, and higher mental processes. *Behav. Sci.*, 1959, **4**, 330–335.

SWETS, J. A. Detection theory and psychophysics: A review. *Psychometrika*, 1961, **26**, 49–63.

TANNER, W. P., JR., & SWETS, J. A. A decision making theory of visual detection. *Psychol. Rev.*, 1954, **61**, 401–409.

(Early publication received July 19, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 131–138

EFFECT OF STRENGTH OF PUNISHMENT FOR "CORRECT" OR "INCORRECT" RESPONSES ON VISUAL DISCRIMINATION PERFORMANCE[1]

GEORGE J. WISCHNER, HARRY FOWLER, AND STEPHEN A. KUSHNICK

University of Pittsburgh

The purpose was to investigate the effect of variation in shock intensity on the acquisition of a black-white discrimination by 56 white rats in a situation involving shock for the correct or incorrect response, a noncorrection procedure, and 2 different training methods (free- vs. forced-choice). Differences in outcome between training methods were negligible. Shock intensity functions showed that errors increased with increasing intensities for shock-right groups, decreased with increasing intensities for shock-wrong groups. All shock-right groups were inferior to no-shock controls, a finding in opposition to the generalization that shock for the correct response facilitates performance.

The results of studies by Muenzinger and his co-workers (Muenzinger, 1934; Muenzinger, Bernstone, & Richards, 1938; Muenzinger & Newcomb, 1936) have been interpreted as indicating that pairing punishment (electric shock) with the *correct* response does not disrupt, but, in fact, facilitates performance. According to Muenzinger, there is a "general sensitizing" factor in punishment which serves to slow down the animal and make it more sensitive to the cues to be discriminated at the choice point of a maze.

Muenzinger's original formulation appears to be generally accepted (e.g., Feldman, 1961; Young, 1961), despite more recent findings which question its generality (Muenzinger & Baxter, 1957; Muenzinger, Brown, Crow, & Powloski, 1952; Muenzinger & Powloski, 1951; Wischner, 1947). Wischner has shown that with a noncorrection procedure, as contrasted with the correction procedure typically employed by Muenzinger, shock for the correct response retards, rather than facilitates discrimination learning. Wischner states that shock has a specific function with respect to the punished response: the animal very quickly acquires avoidance responses to the cues associated with shock.

The results of the more recent studies comparing correction and noncorrection procedures led Muenzinger to modify his original position to one in which it is maintained that shock serves (*a*) to produce avoidance and (*b*) to accelerate learning. Thus, Muenzinger and Powloski (1951) argue that, with a correction procedure, the animal adapts more quickly to the noxious, avoidance-producing quality of shock for the correct response, since selection of the wrong alternative always requires a retracing toward the positive cue with which the shock is associated. As a result, the accelerating or "sensitizing" effect of shock takes place sooner with the correction procedure and a facilitation of performance is readily evidenced.

In addition to the recognition of the significance of methodological variables, such as correction vs. noncorrection, there has been some recognition of the import of factors relating to shock per se. For example, Muenzinger has suggested that only relatively mild shock, when paired with the correct response, may be facilitating. Nevertheless, there has been no systematic study of shock parameters. In relation to the general issue of the effect of shock for the correct response, it would appear that factors such as the intensity, duration, locus, frequency, and order of the shocks administered over training, would constitute a significant class of determining variables.

---

[1] This study was supported in part by Grant No. G-14312 from the National Science Foundation. Grateful acknowledgment is made by the authors to M. Samuels and F. Berlin for their assistance in the conduct of this experiment.

The purpose of this investigation is to assess the effect of different intensities of shock for either the correct or incorrect response on performance in a black-white discrimination entailing a noncorrection procedure. In addition to the shock intensity variable, two training procedures were studied: (*a*) the more usually employed free-choice procedure and (*b*) a forced-choice procedure. The latter procedure was included in order to determine the effect, if any, resulting from equating, among groups, the number and order of shock and food experiences received over training.

METHOD

Subjects.—Fifty-six male albino rats of the Sprague-Dawley strain served as *S*s. These *S*s, which had been selected on the basis of their performance in pretraining tests described below, were approximately 100 days old at the beginning of training. The *S*s were caged individually in a colony room in which the temperature was controlled and the normal day-night cycle reversed through artificial illumination.

Apparatus.—The apparatus was a simple T maze with walls and floor of galvanized sheet metal and an exterior supporting shell of wood. The stem of the apparatus was 26 in. long, and each arm, as measured from the center of the choice point area, 36 in. long. The height of each section of the maze was uniformly 4½ in. and the width, 3¾ in. The first 8 in. of the stem served as a start compartment, being separated from the remaining portion by a guillotine door. The last 12 in. of each arm were also separated from the remaining portion by a guillotine door and served as a goal compartment. Guillotine doors were also positioned in each arm at a distance of 3 in. from the center of the choice point.

The internal sheet metal composition of the maze was provided by two L shaped strips of sheet metal; one L served as one wall and half of the floor of each maze section and the other L as the other wall and half of the floor. The L shaped strips of sheet metal provided two 1½-in. floor surfaces which were separated by a 1-in. gap. Hinged and weighted transparent Plexiglas tops covered the stem and each arm.

A plate of frosted Plexiglas served as the end wall of each goal compartment. A strip of clear Plexiglas (1 in. wide, ½ in. thick, and 3¾ in. long) with a carved food well (¾ in. in diameter and ½ in. deep) was permanently positioned within each goal compartment, directly in front of the frosted plate. A single 10-w. frosted bulb (120-v. source) was positioned outside of each goal compartment, directly behind the frosted plate. These goal-box bulbs, which served as the discriminative visual stimuli, were enclosed by the frosted plate of each goal compartment and an exterior wooden shell painted a flat black. Exterior maze illumination was provided by two 10-w., 120-v. frosted bulbs, one located near the end of each arm on the stem side of the T maze.

A matched-impedence 60-cps ac shock system was used to deliver shock to the *S* in either maze arm. In this system the two L shaped strips of sheet metal forming the sides and floor of the maze were connected across the output of a variable transformer. An additional resistance of .30 megohms was placed in series with *S* which could receive shock when it made contact with both halves of the sheet-metal floor. The narrowness of the maze alley prevented *S* from avoiding shock by running along only one half of the floor surface.

Additional characteristics of this shock circuit were such that it had to be set by *E* before each training trial. Shock was delivered to *S* when it interrupted, in either arm, an infrared photoelectric beam crossing the arm at a point midway between its two guillotine doors. The manual priming feature of the circuit prevented *S* from receiving more than one shock in any one trial, even though it was possible for *S* to interrupt a photoelectric beam several times during a trial. When the shock was delivered, its duration was held constant at .2 sec., as metered through a Hunter timing relay, and its intensity was set at either 45, 60, or 75 v. as measured across the output of the transformer. Preliminary testing for threshold of response to the shock indicated that, with the circuit employed, animals manifested a reaction to shock at approximately 40 v.

Procedure.—The experimental procedure entailed both a pretraining and a training phase. One week prior to the pretraining phase, the *S*s were started, and maintained for the duration of the experiment, on a daily diet of 12 gm. of Purina lab checkers per day, with water available ad libitum. Taming of the *S*s also took place during this week and consisted of a daily handling of each *S* for several minutes.

For both pretraining and training, the *S*s were transported to the experimental room in a wood and hardware cloth carrying cage. The *S*s were kept in individual compartments of the carrying cage throughout the course of the daily training session. Prior to their daily run in the maze, which took place during the dark phase of the *S*'s day-night cycle, all *S*s received approximately 5 min. adaptation to the illumination conditions of the room as provided by the two exterior 10-w. apparatus bulbs, this time being that required by *E* to prepare for the day's trials.

Pretraining was administered to reduce possible position and brightness preferences and to habituate *S* to the apparatus, especially the guillotine doors which were used to prevent retracing. No shock was administered during this phase. Over pretraining, each *S* received a total of 16 forced, reinforced trials, 4 per day, randomly distributed with the restriction that half of the trials were to the left and half to the right with half of each of these sets being to an illuminated goal and half to a dark goal. Forcing was accomplished by lowering one of the guillotine doors at the choice point. The reinforcement provided on each trial was in the form of P. J. Noyes sugar pellets (4 mm., 45 mg.). On the first of the 4 pretraining days, the food pellets were liberally spread throughout the goal box and on subsequent days, the number and locus of pellets were systematically reduced until, on the last day, only two pellets were given and these were in the goal-box food cup. With each forced trial, *S* was permitted to run to the available goal where it was detained until all pellets were consumed or until 5 min. had elapsed. Upon meeting either of these two criteria, or if the *S* failed to run to the forced goal within 5 min., it was removed from the maze and replaced in the carrying cage to await its next trial. On the average, the intertrial interval was about 15 min., the interval being determined by the time required to run a block of *S*s. Those *S*s who consistently traversed the maze and consumed the food pellets within the allotted time were considered for the training phase of the study. This represented about 90% of all *S*s pretrained. It should be noted that those *S*s which were retained showed no reliable brightness or position preferences. Analysis of first training trial choices indicated a 61% selection of the illuminated alley.

Training began on the day following termination of pretraining. During training, reinforcement could be obtained only in the illuminated goal box. Since reinforcement was not obtained in the dark goal box during these training trials, *S*s were detained in that goal box for 10 sec., the approximate time required by *S* to consume two food pellets when in the lighted goal. Otherwise, the conditions of reinforcement, schedule of right or left goal-box illumination, intertrial interval, and general handling procedure were the same for both training and pretraining trials.

For the training phase, the 56 *S*s were randomly assigned to seven groups of 8 *S*s each. There were three shock-right groups (SR), i.e., *S*s received both shock and food reward for a correct response but neither for an incorrect response; three shock-wrong groups (SW), i.e., *S*s received food reward for a correct response but shock for an incorrect one; and a control, no-shock group (NS), i.e., *S*s received only food reward for a correct response. The different shock intensities employed for the three groups within both the SR and SW conditions were 45, 60, and 75 v. The shock-right groups may be designated as SR-45, SR-60, and SR-75; and correspondingly, the shock-wrong groups as SW-45, SW-60, and SW-75.

The six shock groups and the control group were subdivided, with half the *S*s within each group being trained on the typically employed free-choice procedure, and half being trained on a forced-choice procedure. The *S*s run under the forced-choice procedure were permitted a free choice only on the first trial of successive blocks of four trials. The remaining three trials within each four-trial block were forced-choice trials. Forcing was

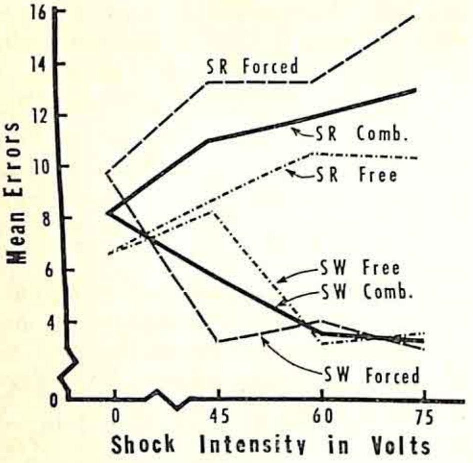


FIG. 1. Group mean errors over 200 training trials as a function of shock intensity.

administered such that, of the four trials within a block, two were to the goal box on the left and two were to the goal box on the right with one trial of each of these sets of two trials being to a lighted goal compartment with reinforcement, and the other being to a dark goal without reinforcement. Thus, after every block of four trials, and at any point in training, occasions of reinforcement and nonreinforcement were equal and balanced over right and left positions for all groups of *S*s run under the forced-choice procedure. Similarly, for those shocked *S*s run under this procedure, the occasions of shock and no-shock were equal and balanced over right and left positions throughout training.

All *S*s received 4 trials per day on the first 6 days of discrimination training and then 8 trials per day thereafter until the end of training. The position of the illuminated, correct goal was varied on these trials in accordance with a predetermined random schedule. Discrimination training was continued until each *S* met a criterion of 15 correct responses out of a total of 16 free choices with the last 8 being correct, or until a total of 200 training trials had been administered. On every trial, each *S* was permitted a maximum of 5 min., from the opening of the start-box door, to make a choice. If *S* did not complete a trial within the allotted time, the trial was recorded as "No-Response" (NR). Actually there were relatively few instances of such trials and these are considered in a subsequent section of the results.

RESULTS

For both the free- and forced-choice groups, only the first trial of every successive block of four was utilized in the major analyses. For the forced-choice groups, this trial was, of course, the only scorable trial. By using only the corresponding first trial of a block of four for the free-choice animals, it was possible to compare performances obtained with the two training procedures. It should be noted, however, that the first-trial data for the free-choice groups are highly representative of their performance over all trials. The results of an analysis of variance involving Trial 1 and averaged total-trial data showed that group mean error differences based on these two measures are insignificant ($F < 1$). Expressed in terms of the correlation between individual error scores based on the Trial 1 and Trial 2–4 data, the product-moment r is .79.

Mean errors over 200 trials for the several groups within both the free- and forced-choice procedures are presented as a function of shock intensity in Fig. 1. The shock intensity functions for both free- and forced-choice procedures appear to be quite similar. In general, mean errors tend to increase with increasing shock intensities for the SR groups, but decrease with increasing intensities for the SW groups. The regularity of this trend is more clearly evident in the heavy, solid line in Fig. 1 which represents the pooled data for corresponding shock-intensity groups of the free- and forced-choice procedures.

Since the no-shock or zero intensity condition is part of the intensity dimension of both the SR and SW conditions, its inclusion in an analysis of variance of the data presented in Fig. 1 precludes the use of a complete factorial analysis of shock conditions (SR vs. SW) and shock intensity.

The data were statistically treated in terms of a 2 × 7 factorial analysis of variance comprising the two training methods and the seven groups employed; i.e., three SR intensity groups, three SW intensity groups, and the NS group. With this analysis, the apparently divergent trends depicted in Fig. 1 for the SW and SR groups, over increasing shock intensities, can be assessed by orthogonal polynomial comparisons.

The results of the analysis of variance, presented in Table 1, show that the between-groups differences are highly reliable ($p < .005$) and that this effect related primarily to the highly reliable differences among the seven shock intensity groups ($p < .001$); the difference between the two training procedures, as well as the Procedure × Shock Condition interaction, is unreliable ($p > .20$ in both cases).

To assess the significance of the differences among the several shock intensities and between shock conditions (SR vs. SW), an orthogonal polynomial comparison was performed assessing an increasing linear trend over increasing SR intensities vs. a decreasing linear trend over increasing SW intensities. One may view this comparison as identical to a linear trend analysis by orthogonal polynomials of all seven groups ordered from SR-75 through NS to SW-75. As such, residual effects after extraction of this linear component, have reference to other trend components, e.g., quadratic, cubic, etc.

As shown in Table 1, this increasing linear SR vs. decreasing linear SW comparison is highly significant ($p < .001$), and indicates not only a reliable difference between SR and SW conditions, but also, as depicted by the solid line in Fig. 1, reliable differences among the several shock intensity groups within both SR and SW conditions. With extraction of this significant linear component, residual trend effects are negligible ($F < 1$). This suggests that the best assumption concerning the nature of the shock intensity functions for both the SR and SW conditions is that they are linear.

The differences obtained above between the SR and SW conditions on the basis of all 200 training trials appear to be representative of any portion or phase of training. Figure 2 shows the course of acquisition of the discrimination over blocks of 40 training trials. The left panel presents the data for *S*s of the free-choice procedure and the right panel for *S*s of the forced-choice procedure. The data for the various SR and SW intensity groups within each training procedure have been pooled in order to reduce the cluster of curves. With both procedures, it is evident that, in comparison with the SW groups, the inferior performance of the SR groups is maintained throughout the 200 training trials.

Analysis of nonresponse (*NR*).—The above analyses do not include NRs made by the *S*s throughout the course of training. Figure 3 shows

TABLE 1

ANALYSIS OF VARIANCE OF GROUP MEAN ERRORS OVER ALL TRAINING TRIALS FOR SR AND SW GROUPS

Source	*df*	*MS*	*F*
Between	13	73.22	2.99**
Procedure (P)	1	34.57	1.41
Condition (C)	6	127.68	5.21***
Linear	(1)	743.14	30.32***
Residual	(5)	4.59	<1
P × C	6	25.20	1.03
Linear	(1)	70.88	2.89
Residual	(5)	16.06	<1
Within	42	24.51	

** $p < .01$.
*** $p < .001$.

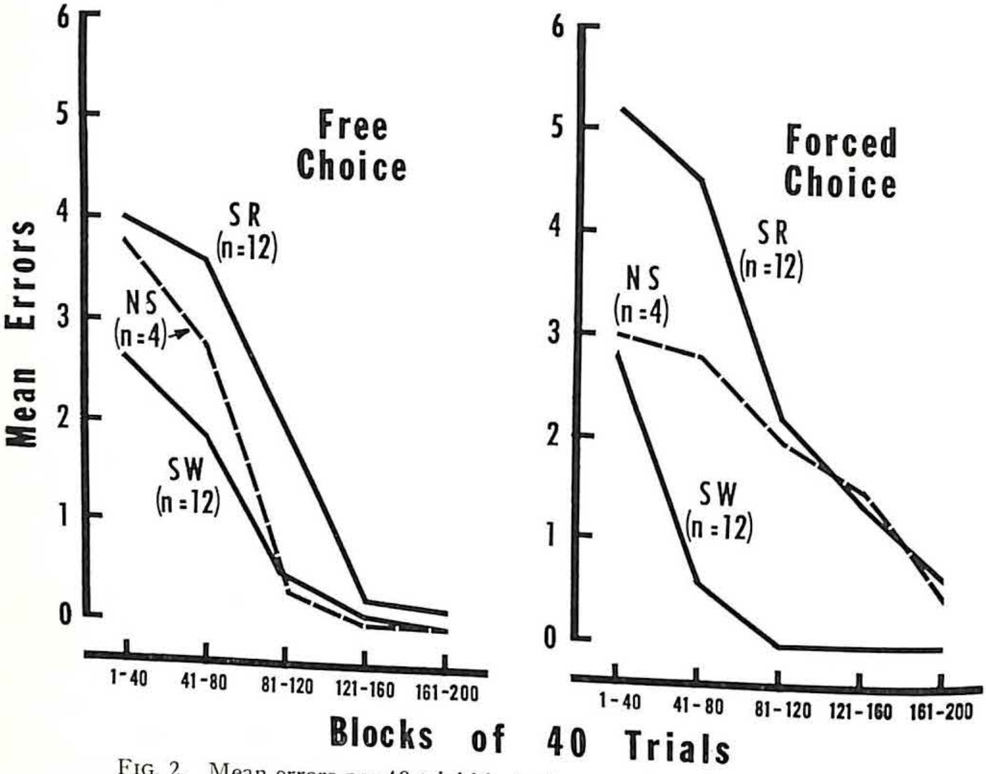


FIG. 2. Mean errors per 40-trial blocks for groups run under free- and forced-choice procedures.

group mean NRs over training as a function of shock intensity. The NR scores for the corresponding groups of the two training procedures have been pooled in view of the absence of a significant difference in the frequency of NRs between training procedures and an insignificant Training Procedure × Shock Condition interaction. Differences in the frequency of NRs among the seven groups shown in Fig. 3 are highly reliable, $F(6,42) = 2.82$, $p < .025$, and trend analysis of this significant effect shows both reliable linear and quadratic components, $F(1,42) = 4.70$, $p < .05$ and $F(1,42) = 11.41$, $p < .005$, respectively. The highly significant quadratic component indicates, as shown in Fig. 3, that the frequency of NRs within both the SR and SW conditions increases with increasing shock intensities; the significant linear component indicates that this effect is somewhat more pronounced for the SR condition.

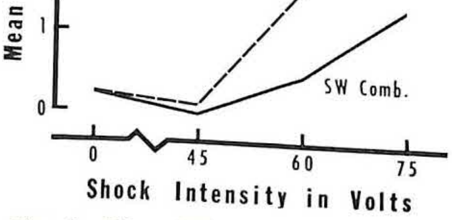


FIG. 3. Mean NRs as a function of shock intensity for the SR and SW conditions with appropriate free- and forced-choice groups combined.

Although this analysis shows systematic differences in frequency of NRs among the several shock groups and between shock conditions, it should be emphasized that these differences seem to be unrelated to the significant effects noted for the earlier error score analyses. Whereas errors increased with increasing SR intensities and decreased with increasing SW intensities, NRs increased with increasing intensities for both the SR and SW conditions. Moreover, it would appear unlikely that the large error differences obtained between the SR and SW conditions could be accounted for by the relatively small absolute, albeit significant, differences in NR noted between these conditions in Fig. 3.

DISCUSSION

In relation to the effect of shock for either the correct or incorrect response, the results of the present study clearly demonstrate the importance of the intensity parameter. Significantly, variation in the intensity of shock associated with the correct response served to vary the extent of performance *retardation*; variation in shock intensity associated with the incorrect response served to vary the extent of performance *facilitation*.

With shock intensities lower than those employed in the present study, some facilitation of performance might possibly be obtained for the SR condition. This appears unlikely in view of the fact that threshold for the kind of shock experience provided the *S*s is obtained at approximately 30 v. (Campbell & Teghtsoonian, 1958). Thus, any facilitation in performance due to shock for the correct response would have to occur within the relatively narrow range from 30–45 v.

The present results are in striking contrast to those obtained by Muenzinger and his co-workers with both correction and noncorrection training procedures. Our findings offer strong support for the view that the primary function of shock is to produce avoidance of the cues associated with it. As Wischner (1947) has indicated, this position does not exclude the possibility that shock may also have the function of slowing down the animal, thus increasing the likelihood of stimulation by and response to the relevant cues, and thereby facilitating performance.

More recently, and primarily on the basis of data obtained in a noncorrection situation, Muenzinger and his co-workers (Muenzinger et al., 1952; Muenzinger & Powloski, 1951) have acknowledged the primary avoidance function of shock. Nevertheless, they (Muenzinger & Powloski, 1951) continue to argue for the accelerating effect of shock for the correct response on the basis of findings, which show that the shock-right condition facilitates performance considerably with a correction procedure but only slightly with a noncorrection procedure. In this connection, the present writers would propose that the "noncorrection" procedure which Muenzinger employed is more comparable to a correction than a noncorrection procedure, as they are typically defined. With Muenzinger's "noncorrection" procedure, *S* was prevented from retracing following entrance into a choice alley, but the random, left-right positioning of the discriminative stimuli from one trial to the next did not take place until *S* made a correct response. Thus, as with a correction procedure, *S* could make several errors or runs to the incorrect goal per "trial," i.e., before the left-right positions of the discriminative stimuli were varied in accordance with their predetermined, random order. It would appear, then, that the inferior performance of the SR groups in the present study, as compared with the relatively better performance of Muenzinger's "noncorrection" shock-right groups, may be related to the degree to which his "noncorrection" method resembles a correction method. This possibility, interestingly enough, has been suggested by Muenzinger and Powloski (1951) in an attempt to reconcile the discrepancy between their results and Wischner's (1947).

Use of either the correction or noncorrection method may underlie the discrepant results obtained with shock for the correct response but this difference does not account for the facilitation of performance which is obtained with one procedure and the retardation obtained with the other. Several explanations of these differences have been offered (e.g., Mowrer, 1950; Prince, 1958; Woodworth & Schlosberg, 1954), but the results of the present study do not contribute to the selection or support of any one of these over another. Rather, the results of this study and those reported more recently by Muenzinger and his associates serve to delimit clearly the generalization, originally proposed by Muenzinger and accepted by many workers in this area, that shock for the correct response facilitates discrimination performance.

REFERENCES

Campbell, B. A., & Teghtsoonian, R. Electrical and behavioral effects of different types of shock stimuli on the rat. *J. comp. physiol. Psychol.*, 1958, **51**, 185–192.

Feldman, S. M. Differential effects of shock in human maze learning. *J. exp. Psychol.*, 1961, **62**, 171–179.

Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

Muenzinger, K. F. Motivation in learning: I. Electric shock for correct responses in the visual discrimination habit. *J. comp. Psychol.*, 1934, **17**, 439–448.

Muenzinger, K. F., & Baxter, L. F. The effect of training to approach vs. training to escape from electric shock upon subsequent discrimination learning. *J. comp. physiol. Psychol.*, 1957, **50**, 252–257.

Muenzinger, K. F., Bernstone, A. H., & Richards, L. Motivation in learning: VIII. Equal amounts of electric shock for right and wrong responses in a visual discrimination habit. *J. comp. Psychol.*, 1938, **26**, 177–186.

Muenzinger, K. F., Brown, W. C., Crow, W. J., & Powloski, R. F. Motivation in learning: XI. An analysis of electric shock for correct responses into the avoidance and accelerating components. *J. exp. Psychol.*, 1952, **43**, 115–119.

Muenzinger, K. F., & Newcomb, H. Motivation in learning. V. The relative effectiveness of jumping a gap and crossing an electric grid in a visual discrimination habit. *J. comp. Psychol.*, 1936, **21**, 95–104.

Muenzinger, K. F., & Powloski, R. F. Motivation in learning: X. Comparison of electric shock for correct turns in a corrective and non-corrective situation. *J. exp. Psychol.*, 1951, **42**, 118–124.

Prince, A. I., Jr. Effect of punishment on visual discrimination learning. *J. exp. Psychol.*, 1956, **52**, 381–385.

Wischner, G. J. The effect of punishment on discrimination learning in a noncorrection situation. *J. exp. Psychol.*, 1947, **37**, 271–284.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

Young, P. T. *Motivation and emotion: A survey of the determinants of human and animal activity.* New York: Wiley, 1961.

(Received February 12, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 139–147

LONG-TERM EFFECTS OF MISINFORMATIVE FEEDBACK UPON CONCEPT IDENTIFICATION[1]

LYLE E. BOURNE, JR.

University of Utah

The purposes of this study were to investigate the long-term effects of misinformative feedback (MF) upon performance in a concept identification problem (Task 1) and the transferability of optimal solution to a 2nd problem (Task 2). The results indicated that as amount of MF increased in Task 1, Ss made more errors and fewer Ss reached optimal solution. Transitory probability matching was observed. All Ss attained optimal solution to Task 2, at a rate, however, which was a function of Task 1 MF. There were no marked recency effects and little evidence that Ss attempted to respond with the category indicated by the feedback signal accompanying each stimulus on its last presentation.

Goodnow and Postman (1955), Goodnow (1955), and Pishkin (1960) have investigated the effect of misinformative feedback (MF) upon the performance of human *S*s in concept learning tasks. In the procedures of these experiments, MF was introduced by presenting feedback signals, ordinarily used to indicate the correctness or incorrectness of *S*'s response to the given stimulus, which were misleading on a percentage of trials. That is, the signal occasionally indicated to *S* that he had responded correctly or had responded incorrectly when in fact the reverse was true. Results of these studies were essentially the same as those of experiments concerned with guessing behavior (e.g., Grant, Hake, & Hornseth, 1951); i.e., in the later trials there was close agreement between the probability of an MF trial and the probability of an incorrect response. Goodnow and Postman reported that *S*'s responses did not appear to be affected by characteristics of the stimulus sequence, as is the case in studies of guessing behavior (Anderson, 1960; Nicks, 1959), but concluded that the overall similarity of results attests to the generality of probability learning.

In the data reported by Goodnow and Postman, by Goodnow, and by Pishkin there is a tendency on the part of experimental groups with low percentages of MF trials to produce errors at a rate below that predicted by a probability matching hypothesis. For example, the "90:10 group" (10% MF) of Goodnow and Postman consistently produced an average of less than 10% errors over the last six 10-trial blocks of the experiment. Such evidence suggests that, under certain circumstances, *S*s in concept learning situations may seek and attain a solution more rational than probability matching; i.e., a solution which maximizes the number of "hits" or trials upon which response and feedback signal correspond. It is plausible that probability matching shown in the above mentioned studies represented a transitory stage in performance and that all *S*s, if given a larger number of trials, would have adopted the optimal solution which requires consistent responding to the relevant stimulus dimension.

---

[1] This research was supported in part by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service. The author is grateful to F. Restle for his numerous suggestions on the initial phases of the experiment.

Results of a preliminary study, the procedure of which was patterned after that of Pishkin, supplied convincing evidence of the transitory nature of probability matching in concept identification. After extensive training on a problem with MF on 10% of the trials, all *S*s reached optimal solution. In light of these data, the present experiment was undertaken. Its main purpose was to discover if, under different conditions of MF percentage, *S*s would reach an optimal solution in a concept identification task. Furthermore, the experiment could supply additional evidence with respect to the cues mediating probability learning and the attainment of an optimal solution in concept problems. Secondarily, the study was designed to find out whether an attained optimal solution is transferable to a similar problem which involves a different degree of MF. To collect evidence on this question, all groups were shifted without instruction to a second concept identification problem after lengthy practice on the first.

METHOD

Task and procedure.—The task was essentially that used by Pishkin (1960). The *S* was required to press one of two buttons, neither of which bore a label, in response to each of a series of geometric stimulus patterns. In all problems, the patterns, which were varied within three two-level dimensions, were projected for viewing on a milk glass screen. At the outset, *S*s were given standard formal instructions describing the stimulus dimensions, and how they might be used as a basis for responding, and pertinent apparatus, and how it was to be operated. The *S*s were told that the correct response on any trial depended on the attributes of the stimulus patterns and nothing else. They were led to believe that errorless performance was possible. Presentation of stimuli, presentation and control of informative feedback (provided by signal lamps located directly above *S*'s two response buttons), control of the overall trial interval (5 sec.), and recording of responses were accomplished automatically with the concept identification apparatus described by Pishkin. In all cases, the feedback light was given immediately after the response and remained for 1 sec. In all problems, *S* responded to stimulus patterns at a self-determined rate.

For each of two problems, a different stripfilmed series of stimulus patterns was used. In Problem A, stimulus dimensions were orientation (upright or tilted figures), form (triangle or square), and color (red or green); and, in Problem B, the dimensions were vertical position (figures at the top or bottom of the viewing screen), size (large or small), and horizontal position (figures on the right or left side of the screen). One dimension, form in Problem A and horizontal position in Problem B, was relevant in the sense that responses based upon it would yield the maximum number of hits attainable under the prevailing condition of MF. The remaining dimensions were irrelevant, responses based upon them producing a chance number (50%) of correct responses.

Under some experimental conditions, *S* was presented with MF, defined as an erroneous feedback signal designating the incorrect response button as correct, on a certain percentage of trials. The scheduling of MF trials was random with two restrictions: (*a*) the total number of misinformed trials was consistent with MF percentages employed in the experimental design, and (*b*) the percentage of MF was approximately the same within each block of 50 trials. Because the stimuli varied in three binary dimensions, there were eight distinctly different patterns. The order in which these were presented was random except that each pattern appeared equally often in every 200 trials and that the same pattern never appeared on successive trials.

Subjects and design.—The *S*s were 80 students from introductory psychology courses who were assigned by order of appearance to the eight treatment conditions of the experiment. Basically, the design was a 4 × 2 factorial with four percentages of MF, 0, 10, 20, and 30, and two different problems. The design was complicated, however, in two ways: (*a*) *S*s served for 6 successive days and were given 200 trials each day, and (*b*) both the MF percentage and the stimulus dimensions were changed for all *S*s on Day 4; if *S* had been performing on Problem A, Problem B was substituted for it and vice versa. All *S*s performed with 15% MF from Day 4 through Day 6. If, during any daily session, *S* made 60 consecutive responses based on the relevant dimension, he was released from the experiment for that day.

RESULTS

The main dependent variable was number of errors made with respect to the relevant stimulus dimension. Thus, for example, all Category 1 responses to squares and Category 2 responses to triangles in Problem A were scored correct, regardless of the accompanying feedback signal. Analyses of variance were performed separately on errors per block of 20 trials made during the first 3 days (Task 1) and during the last 3 days (Task 2) of the experiment. These analyses are summarized in Table 1, and the data are plotted in Fig. 1.

Task 1

Analysis of variance.—Total errors made in the four MF conditions on Task 1 differed reliably. As is obvious from an inspection of Fig. 1, errors increased with amount of MF. In addition, a significant practice effect, within and across days, was evident. The most important finding, however, is the statistical reliability of the MF × Days interaction; clearly the trend in performance across days differed for the various MF conditions. The significance of this interaction is attributable to the fact that the four groups approached their (possibly different) asymptotes at different rates. The *S*s in the 0% MF condition reached problem solution; i.e., consistent responding on the basis of the relevant dimension, within 100 trials, although a few additional errors were made by these *S*s during the first block of 20 trials on each of the succeeding 2 days. Similar warm-up effect is noticeable in the records of all groups. The performances of the 10%, 20%, and 30% MF groups display probability matching; i.e., group functions tended first toward a level of errors which would be predicted if all *S*s were

TABLE 1

ANALYSIS OF VARIANCE ON ERRORS

Source	*df*	Days 1–3		Days 4–6	
		MS	*F*	*MS*	*F*
Misinformative feedback (MF)	3	4037.32	31.25**	20.20	
Problems (P)	1	322.66	2.50	185.37	5.01*
MF × P	3	141.03		31.10	
*S*s/Groups	72	162.95	26.28**	36.97	10.16**
Days (D)	2	1150.63	185.60**	519.43	142.69**
Trials/D	27	79.30	12.79**	61.49	16.89**
MF × D	6	135.11	21.79**	22.05	6.06**
Linear × Linear	1	79.39	12.80**	2.00	
Linear × Quadratic	1	163.52	26.38**	24.01	6.60*
Quadratic × Linear	1	500.64	80.75**	53.06	14.58**
Quadratic × Quadratic	1	42.00	6.77**	11.51	3.17
Cubic × Linear	1	7.38	1.19	28.58	7.86**
Cubic × Quadratic	1	24.73	3.96*	13.14	3.61
P × D	2	0.77		86.08	23.65**
MF × P × D	6	39.82	6.42**	7.15	1.96
Pooled residual	2277	6.20		3.64	

* $p < .05$.
** $p < .01$.

responding incorrectly to the relevant dimension on a percentage of trials equal to the MF percentage. No group remained at this level throughout the 3 days of practice, however. By the end of the third day, 10% MF *S*s had attained the same asymptotic performance as *S*s in the 0% MF condition. Trends toward optimal performance were evident in the records of *S*s in the 20% and 30% MF groups.

Classes of performance.—The error scores of all *S*s on the last three blocks of 20 trials of each day were divided into three mutually exclusive and exhaustive classes. Included in Class 1 were *S*s who made no more than two errors during this series of trials; in Class 2 were those whose error rate approximated the rate of MF (actually, any *S* whose error rate did not differ from the MF percentage by more than 10% within each block of 20 trials was included in this group); in Class 3 were those whose error rate approximated 50% or chance performance. The *S*s in Class 3, it can be concluded, were responding either at random or, more probably, on the basis of one of the irrelevant stimulus dimensions. The frequency with which each class was found within each MF condition on each day of practice is shown in Table 2 along with corresponding mean percentages of errors.

TABLE 2

Analysis of Performance Classes

Day	MF Condition (%)	Class 1	Class 2		Class 3	
		N	*N*	% Errors	*N*	% Errors
1	10	14	2 (14)	13	4	46
	20	9	2 (10)	27	9	48
	30	5	4 (7)	32	11	54
2	10	18	2 (18)	14	0	
	20	13	2 (11)	16	5	49
	30	9	5 (10)	27	6	45
3	10	20	0 (8)		0	
	20	15	3 (7)	13	2	42
	30	7	7 (9)	32	6	54
4	0	15	4 (16)	20	1	45
	10	18	2 (12)	18	0	
	20	17	3 (11)	23	0	
	30	16	2 (7)	16	2	54
5	0	18	0 (10)		2	41
	10	20	0 (10)		0	
	20	20	0 (5)		0	
	30	20	0 (9)		0	
6	0	20	0 (5)		0	
	10	20	0 (8)		0	
	20	20	0 (4)		0	
	30	20	0 (3)		0	

Note.—The number of *S*s displaying Class 2 performance during *any* block of 60 consecutive trials on each day is shown in parentheses.

The frequencies in Class 2 are small at all stages of practice, however, the number of Class 2 performances over *any* block of 60 consecutive trials on each day (shown in parentheses under Class 2, Table 2) is large and suggests that probability matching was a transitory but identifiable stage of performance for most *S*s. The group functions of Fig. 1 are a somewhat misleading description of the learning processes taking place. Even when group functions approximate probability matching, the behavior of most *S*s within groups cannot be similarly described.

Event sequences.—There is sufficient evidence in previous results (e.g., Goodnow, 1955; Hake & Hyman, 1953; Nicks, 1959) to conclude that individual choices in guessing situations depend upon the short event sequences preceding each response as well as the overall proportions of alternatives. Thus, an analysis of category responses after various lengths of event-runs in the present experiment was undertaken.

It is conceivable though unlikely that *S* might base his category response on the preceding correct

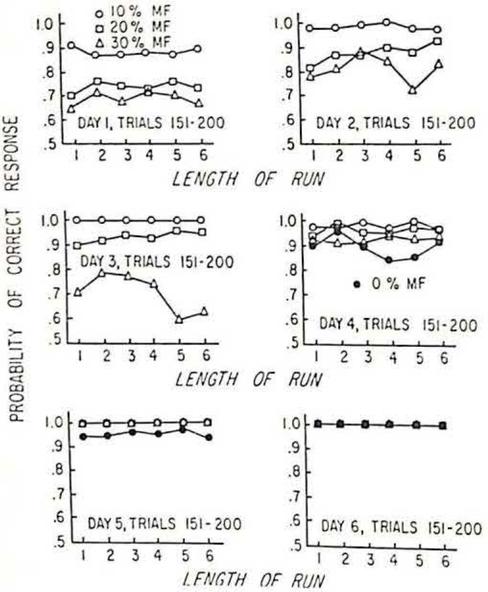


FIG. 2. Probability of a correct response to either level of the relevant dimension as a function of the length of the preceding run of non-MF trials.

response or responses as signaled by the feedback lights, without attending to stimulus patterns. Mean percentages of Category 1 responses after one, two, three, and four consecutive Category 1 signals and of Category 2 responses after identical numbers of Category 2 signals were tallied. There were no consistent changes in response probability as a function of feedback runs over the first 100 trials on Day 1. Since instructions minimized the usefulness of this basis of responding and since all *S*s eventually adopted a mode of response based on stimulus characteristics, the analysis was not carried further.

Figure 2 summarizes part of an analysis of responses to either level of the relevant dimension as a function of runs of non-MF or correct feedback trials. These computations were made for each successive block of 50 trials on each day separately to show any systematic changes in response probability as *S*s progressed from chance to asymptotic performance. The values plotted in Fig. 2 give the percentage of times *S* responded with the proper category to one of the two levels of the relevant dimension as a function of the number of preceding consecutive non-MF trials for the last 50 trials of each day of practice. The results suggest no strong effect attributable to run length in either the 10% or 20% MF conditions before, during, or after the time of apparent probability matching with the possible exception of a slight rise in percentage of correct responses from one to two consecutive non-MF trials. By contrast, Edwards (1961) reported negative recency effects in performance on early trials of a two-alternative learning task which disappeared later in practice. As practice progressed, all percentages for the 10% MF condition approached 100. The performance of 30% MF *S*s was more in accord with that usually found in guessing games; there was a tendency for percentages to decrease as the run of non-MF trials increased beyond 2, particularly on the last 150 trials of Day 3.

Previous feedback matches.—Because in any problem there were only eight distinctly different stimulus patterns, *S* might attempt to remember the category indicated for each by the feedback signals. Incidences of this type reported by Goodnow and Postman (1955) and observations in the present experiment suggested that an analysis be performed to determine the extent to which *S*s' responses to each stimulus were a function of the category indicated to be correct on the last appearance of that stimulus in the sequence. Note that, if *S* does recall and does make the response indicated to be correct for each stimulus on its last presentation; i.e., matches the previous feedback signal, his percentage of errors will be the prevailing MF percentage exactly. Thus, it is conceivable that the observed transitory probability matching in this experiment is the result of an attempt by *S* to match previous feedback.

The percentage of trials on which *S*'s response to each stimulus was the same as the feedback signal accompanying that stimulus on its last presentation (percent PFM) was computed separately for successive blocks of 100 trials in each MF condition. The only evidence of successful matching on previous feedback appeared on Days 2 and 3 in the performance of 30% MF *S*s. In all other conditions, obtained percent PFM values were less than the non-MF percentages. The analysis, of course, does not demonstrate the complete lack of matching. The percent PFM values of some *S*s in several conditions were considerably above the prevailing non-MF percentages. However, the number of such *S*s and the success they had were too small to produce any clear effect in this experiment.

Task 2

Analysis of variance.—The overall *F* test for differences in total errors in Task 2 among the MF groups was not significant. It is noteworthy that the 0% MF condition produced the largest mean number of errors, 36.4, as compared to 24.6, 23.7, and 25.1 for the 10%, 20%, and 30% MF groups, respectively. The 0% MF group differed reliably from the remaining groups ($p < .01$) according to Duncan's range test. As in Task 1, reliable practice effects within and across days were recorded. Of more interest is the MF × Days interaction; the trends in performance exhibited by the four groups differed significantly. Figure 1 and the significant Quadratic × Linear and Cubic × Linear orthogonal components of this interaction indicate that, although all groups reached optimal solution to the second problem, they progressed toward this asymptote at different rates. In particular, the 0% MF group approached this level more slowly than any other, the difference being most obvious on Days 4 and 5. There is little variation among the rates of the remaining groups.

Classes of performance.—Table 2 shows that the majority of *S*s did make errors at a rate consistent with the MF percentage (15%) at some stage of practice but that no *S* continued to make errors at this rate for the duration of the experiment.

Events sequence.—An analysis of percentage of correct responses as a function of number of preceding non-MF trials was performed on Task 2 data. There was no observable length of run effect in the performance of 10%, 20%, and 30% MF *S*s; their percentages were homogeneous within blocks of trials and progressed steadily toward 100. The 0% MF group showed observable recency effects on Day 4 which gradually decreased as optimal solution was attained.

Previous feedback matches.—There was no identifiable tendency for *S*s to respond to each individual stimulus pattern with the category indicated to be correct on its preceding appearance in the sequence. Because all *S*s eventually adopted an optimal solution in this task, percent PFM values for all groups approximated 85 at the end of practice.

DISCUSSION

While results of this experiment do not contradict those reported by Goodnow and Postman (1955) and by Pishkin (1960), they are not in strict agreement with the conclusions reached in these studies. Most *S*s, at some time during their search for a lawful solution to the concept problem, tended to commit errors at a rate commensurate with the percentage of MF trials. However, probability matching in the present experiment was clearly transitory. All *S*s in the 10% MF condition and many in the 20% and 30% conditions eventually attained an optimal solution in Task 1. In Task 2, all 80 *S*s reached the optimal solution.

Several analyses were performed to determine what cues mediated probability matching and the eventual attainment of optimal solution. In guessing game studies, matching is partially the result of guessing tendencies related to the structure of event sequences. The reasonable supposition that similar cues affected the behavior of *S*s in the present task appears to be untenable, although there is evidence: (*a*) that the probability of a correct response was higher after two consecutive non-MF trials than after one such trial, and (*b*) of a negative recency effect in the performance of 30% MF *S*s on Task 1 and of 0% MF *S*s on Task 2. In other circumstances, probabilities of correct response were homogeneous across lengths of non-MF trial runs. A similar observation of performance in a different but equally complex learning task has been reported by Voss, Thompson, and Keegan (1959).

It may be, as Goodnow (1955) has argued, that recency effects fail to occur in tasks which involve trial-to-trial variations in stimulus material and toward which *S* is instructed to assume a problem solving orientation. Since *S* is not set to gamble or guess in such tasks, the guessing habits he might otherwise employ, which are presumed to produce recency effects, do not form a significant part of his response repertoire. Furthermore, the short run patterns which *S* might use as a basis of response choice are obscured by multidimensional stimulus variations. Probability matching observed in the present experiment may be entirely a function of *S*'s search for a cue or set of cues which, in conjunction with the relevant dimension, would permit him to match the feedback signal on every trial. Alternatively, since each of the eight different stimulus patterns was displayed many times during each daily session, *S* may use as a basis of response the category recalled to be correct for the previous presentation of the current stimulus pattern. This type of behavior was identifiable in the performance of individual *S*s and in the overall performance of the 30% MF group only. Furthermore, such behavior was eliminated in every case in favor of the less complex and more rational tendency to respond consistently to the relevant dimension. The fact that responses to previous feedback were most evident in the performance of 30% MF *S*s suggests that attempts to remember each individual stimulus pattern and its category are only undertaken in situations in which the relevant dimension is not clearly differentiated by feedback from the irrelevant dimensions and/or in which the number of hits, even under optimal solution, is close to chance. Two experiments presently in progress, in which: (*a*) MF percentages greater than 30 are explored, and (*b*) MF percentage is varied among the eight different stimulus patterns, may provide additional evidence with respect to this hypothesis.

The data analyses indicate that no single and definable set of cues mediates either probability matching or the attainment of an optimal solution. Furthermore, they suggest that any attempt to present a single description of performance by pooling data from all *S*s would only obfuscate the pronounced individual differences and is of questionable scientific validity.

Regardless of the reasons for or interpretation of *S*'s performance, it is clear that he learns that more hits can be attained by consistently responding to the relevant dimension than by any other means. Furthermore, the optimal solution is transferable to a new task with a different percentage of MF. Those groups which were exposed to MF during Task 1 showed performance on Task 2 superior to that of the 0% MF group; 0% MF *S*s also displayed more probability matching on Task 2 than any other group, as expected. The lack of any essential difference among the 10%, 20%, and 30% MF conditions in Task 2 performance supports the conclusion that, within the limits of this experiment, the general mode of performance characterized by consistent responding to the relevant dimension can be acquired under any MF percentage. This outcome is similar to the effects of prior experience with certain event sequences upon performance on two-alternative learning situations reported by Goodnow and Pettigrew (1955) and may be interpreted as a reflection of the lack of opportunity for *S*s in the 0% MF condition to learn the optimal response strategy for Task 2.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

EDWARDS, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

GOODNOW, J. J. Response sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 624–630.

GOODNOW, J. J., & PETTIGREW, T. F. Effect of prior patterns of experience upon strategies and learning sets. *J. exp. Psychol.*, 1955, **49**, 381–389.

GOODNOW, J. J., & POSTMAN, L. Probability learning in a problem-solving situation. *J. exp. Psychol.*, 1955, **49**, 16–22.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

NICKS, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

PISHKIN, V. Effects of probability of misinformation and number of irrelevant dimensions upon concept identification. *J. exp. Psychol.*, 1960, **59**, 371–378.

VOSS, J. F., THOMPSON, C. P., & KEEGAN, J. H. Acquisition of probabilistic paired associates as a function of S-R_1, S-R_2 probability. *J. exp. Psychol.*, 1959, **58**, 390–399.

(Received February 15, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 148–155

EXTINCTION PERFORMANCE FOLLOWING DISCRIMINATION TRAINING[1]

DAVID BIRCH, JAMES K. ALLISON, AND ROBERT F. HOUSE

University of Michigan

The resistance to extinction of the positive stimulus in discrimination learning was evaluated in 2 experiments using rats and single stimulus presentation on an elevated runway. In Experiment I, discrimination groups received reinforcement to white and nonreinforcement to gray (WG) or black (WB). Reference groups received all trials to white under 100% (W+ and w+ which differ in number of trials) or 50% (W±) reinforcement. Both mean log speed of approach and parameter estimates from the extinction curves showed (Tukey tests) the extinction performance of the discrimination groups as more similar to W+ and w+ than to W±. Experiment II yielded comparable results for WG and WB for both the approach and an additional goal platform measure.

In simple discrimination learning, responses to the positive and negative stimuli are provided 100% and 0% reinforcement, respectively. This procedure may not, however, result in continuous reinforcement of responses to all aspects of the positive stimulus. If generalization between the positive and negative stimuli occurs, the possibility exists of noncontinuous reinforcement for the positive stimulus. Aspects of the positive stimulus which are similar to aspects of the negative stimulus would then be associated with a sequence of both reinforced and nonreinforced trials during the learning period. Such a sequence corresponds theoretically to that of partial reinforcement and hence might be expected to produce the resistance to extinction appropriate to such a schedule.

The hypothesis following from this analysis is that responses to the positive stimulus of a discrimination will be more resistant to extinction than responses to a continuously reinforced stimulus. Pertinent theoretical and experimental research is reported by Skinner (1938), Wickens and Snide (1955), Amsel (1958), and Jenkins (1961), who also summarize the earlier work. Neither Skinner nor Wickens and Snide obtained results giving strong support to the hypothesis. However, Jenkins (1961), employing the key peck response of the pigeon, reports evidence favorable to the hypothesis.

The present study is directed to a further test of the hypothesis. Five groups of rats were used in the first experiment and training and testing was carried out with methods of single stimulus presentation on an elevated runway. A white goal platform was present on all reinforced trials for all groups. Two groups (W+ and w+) differed in total number of acquisition trials and only received reinforced trials to the white goal platform. Three additional groups received half reinforced and half nonreinforced trials. The white goal platform was present on reinforced trials for all groups, but one group (W±) received its nonreinforcement on white, a second group (WG) on gray, and a third group (WB) on black. Comparisons of the extinction performance of the WG and WB discrimination groups to the W+, w+, and W± reference groups were of primary interest.

---

[1] These experiments were supported by National Science Foundation Grant No. G9779 to the senior author.

EXPERIMENT I

Method

Subjects.—The *Ss* were 70 male rats obtained from the Windsor Biology Gardens, Bloomington, Indiana. The *Ss* were experimentally naive and approximately 90 days old at the beginning of the experiment. Seven *Ss* were discarded for refusal to run and 3 for apparatus failure during the experiment.

Apparatus.—The elevated runway described previously by Birch, Ison, and Sperling (1960) was used. This apparatus consists of two elevated platforms enclosed in a cubical compartment, 8 ft. to a side and painted flat black. A T shaped partition, also 8 ft. high, divides the compartment into an approach platform section, a goal platform section, and an entranceway for *E*.

The approach platform is 12 in. long and 3 in. wide and placed 26 in. from the floor. The end of the approach platform extends into the goal platform section through a doorway in the partition. The goal platform, 12 in. long and 4 in. wide, is placed in line with the starting platform but 1 in. lower and separated from it by a 1-in. gap. A small unpainted aluminum foodcup could be attached to the terminal end of the goal platform. The approach platform is painted black while either a black, gray, or white goal platform could be presented by an arrangement of snap fasteners.

A skylight of translucent plastic, 2 ft. square, is located in the cube directly over the goal platform. A 150-w. bulb in circuit with a rheostat set to give a constant illumination of 5.29 ft-L to the goal platform is attached to the ceiling of the outer room directly over the skylight. The approach platform is illuminated by a 7-w. bulb attached to the partition 3 ft. above. An infrared light beam, obtained from two photoelectric cells placed outside the cube, is directed 1 in. over the approach and 1 in. from the partition. When *S* breaks this beam the light in the approach platform section goes out, the light above the goal platform skylight goes on and illuminates the goal platform, and a Standard Electric clock is started. When *S* steps off the approach platform, a normally off microswitch under the platform is released and stops the clock. Latency of response therefore is measured from the onset of illumination in the goal platform section until *S* removes its weight from the approach platform. A black door is lowered by *E* from the outer room to prevent retracing.

Procedure.—Preliminary training was carried out over 14 days. The *Ss* were handled 5 min. daily for 8 days, pretrained on a natural pine elevated runway during the next 3 days, and accustomed to the experimental apparatus on the final 3 days. A 22-hr. food maintenance schedule was imposed during preliminary training and the experiment proper by providing *S* with access to regular laboratory food for 1 hr. each day. The *Ss* were housed individually and water was always available.

Five groups (W+, W±, WG, WB, and w+) differed in their acquisition conditions. Group W+ was presented with the white platform only for 36 trials and was 100% reinforced for approaching it. Groups W±, WG, and WB received 18 reinforced trials to white and 18 nonreinforced trials. Group W± was given nonreinforcement to white, Group WG to gray, and Group WB to black. Group w+, a 100% reinforcement group like W+, received 18 reinforced trials to the white platform and was included to control for the total number of reinforced trials.

Groups W+, W±, WG, and WB were given 14 acquisition trials on Day 1, 14 on Day 2, and 8 on Day 3. The pattern of reinforced and nonreinforced trials for Groups W±, WG, and WB was − + − + − − + + + − + + − − on Day 1, the opposite, + − + − + + − − − + − − + +, on Day 2, and − + − − + + − + on Day 3. Group w+ was provided 10 acquisition trials on Day 1 and 8 on Day 2. On the extinction day, Day 3 for Groups W+, W±, WG, and WB, and Day 2 for Group w+, all groups were presented with 8 acquisition trials followed by 19 nonreinforced extinction trials to the white platform.

A single .05-gm. food pellet baited the foodcup on reinforced trials and *S* was detained on the goal platform for 15 sec. with no foodcup present on nonreinforced trials. A minimum intertrial interval of 5 min. was maintained by running *Ss* in squads of 3. The transition from acquisition to extinction occurred without change in the intertrial interval. Two *Es*, JKA and RFH, were each responsible for one-half of the *Ss* within each group.

TABLE 1

MEANS AND *SD*s OF MEASURES APPROPRIATE TO EVALUATION OF EXTINCTION PERFORMANCE FOR APPROACH SPEED

Measure		Group				
		W±	WG	WB	w+	W+
Extinction speed	Mean	274	174	206	177	194
	SD	63	50	47	44	45
Terminal acquisition speed	Mean	310	227	281	258	323
	SD	47	78	46	69	44
C	Mean	−.11	.37	.57	.48	1.01
	SD	.61	.55	.72	.78	.73
B	Mean	−5.32	−3.12	−5.14	−4.84	−6.98
	SD	4.15	3.74	2.75	4.94	3.30
A	Mean	277	164	190	164	166
	SD	71	57	60	48	51
$(2.432)A - B^2$	Mean		375	429	352	345
	SD		132	141	89	108
$(1.320)A - B^2$	Mean	320	192	217	169	
	SD	114	72	79	50	

Results

Extinction data for the five groups are plotted in terms of log response speed in Fig. 1. Inspection of the curves suggests that the two reference groups (W± and W+) are appreciably different and that the other three groups (WG, WB, and w+) are more similar to the W+ group than to the W± group. Table 1 presents the mean log response speed during extinction for each group. Simple analysis of variance yields $F(4,55) = 7.156$ which is significant beyond the .001 level. Tukey tests (Scheffé, 1959), using the 95% confidence interval on all possible pairs of comparisons among the five means, shows Group W± significantly different from each of the other four groups, but no comparison among the W+, w+, WB, and WG groups yields a significant difference.

Groups WB and WG both attained satisfactory discrimination performance as measured by overall differences in log response speed to the positive and negative stimuli. Both groups respond faster to the positive stimulus; $t(11) = 5.52$, $p < .001$ for WB; and $t(11) = 2.65$, $p < .05$ for WG.

Interpretation of the obtained differences in performance during extinction among the five groups is complicated by the fact that the groups did not enter extinction with a common terminal acquisition speed. Table 1 contains the mean response speed on the last four presentations of the W stimulus for the W±, WG, and WB groups and on the four corresponding trials for the W+ and w+ groups. Simple analysis of variance on these data shows significant differences among the groups, $F(4,55) = 4.76$, $p < .005$, and Tukey tests using the 95% confidence interval show significant differences between WG and W± and between

WG and W+. In addition, the correlation between terminal acquisition log response speed and mean extinction log response speed is $r = .67$ for all groups combined and $r = .92, .78, .65, .55$, and $.74$ for groups W±, W+, w+, WB, and WG, respectively.

The following procedures were carried out in an attempt to obtain information concerning the similarity of the extinction performance of the five groups that is unconfounded by differences in terminal acquisition level. It is assumed that the course of extinction for all 60 *S*s, and thus for the five groups, is the same (i.e., that the same curve would fit the data for all five groups). Differences in terminal acquisition level are interpreted simply as differences in initial points for extinction in reference to a common extinction curve and the procedure of the analysis allows for translation of each individual extinction curve parallel to the abscissa. The null hypothesis is that all differences among individual and group extinction curves will disappear if an appropriate lateral shift of each curve is carried out. That is, obtained parameter values for the fitted function will then be the same for all curves.

It is assumed that the function $\log Y = a + bX + cX^2$, where Y = response speed and X = number of extinction trials, is adequate to represent the extinction data. This function assumes that extinction is exponential in form. Figure 1 indicates that log response speed is not linear in relation to the number of extinction trials, so a second degree term is included in the exponent to accommodate the curvature. The function $\log Y = a + bX + cX^2$ may be fit to data by the methods of orthogonal polynomials (Anderson & Houseman, 1942). This was done for the last 18 trials of extinction for each of the 60 *S*s.

Each *S*'s data were fitted by $\log Y_i = A_i + B_iX + C_iX^2$ where X runs from 1 to 18. Under the assumption of a single common extinction function for all *S*s, however, the data for individual, i, should be fit by $\log Y_i = a_i + b_iX_i + c_iX_i^2$, where the X_i are counted for each individual in relation to the common extinction function. The function, $\log Y_i = A_i + B_iX + C_iX^2$, can be transformed into the desired form by letting $X = X_i + K_i$, where K_i is the unknown amount by which X_i has been shifted in order to carry out the curve fitting with the uniform set of X values.

By substitution,

$$\begin{aligned}\log Y_i &= A_i+B_i(X_i+K_i)+C_i(X_i+K_i)^2\\ &= (A_i+B_iK_i+C_iK_i^2)\\ &\qquad +(B_i+2C_iK_i)X_i+(C_i)X_i^2.\end{aligned}$$

Thus, $a_i = A_i + B_iK_i + C_iK_i^2$, $b_i = B_i + 2C_iK_i$, and $c_i = C_i$; where

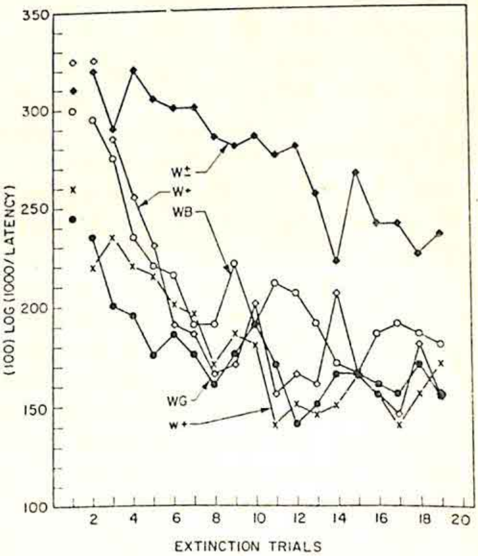


FIG. 1. Extinction performance of the discrimination and control groups with the approach measure.

A_i, B_i, and C_i have numerical values from the curve fitting process and a_i, b_i, c_i, and K_i are unknown.

It will be noted that $c_i = C_i$ so that questions concerning the c parameter may be evaluated directly from the obtained C values. In addition, by eliminating K_i in the two remaining equations, $4c_ia_i - b_i^2 = 4C_iA_i - B_i^2$ is obtained making tests of this composite of the a, b, and c parameters possible. In the case of the present experiment the five groups were compared with respect to mean C values and mean $(4CA - B^2)$ values since the null hypothesis proposes no differences among groups for these quantities. Table 1 also summarizes the means and *SD*s of the parameter values for the five groups.

An indication of the adequacy with which the function $\log Y_i = A_i + B_iX + C_iX^2$ fits the data for the 60 *S*s may be obtained from the ratio of the *SS* for the linear and quadratic components to the total *SS*. These values range from a low of .002 to a high of .787 with a median of .364. Medians of .422, .365, .348, .326, and .414 were found for groups W±, WG, WB, w+, and W+, respectively. These values indicate the data for the five groups are fitted with moderate but comparable success.

Simple analysis of variance comparing the five groups with respect to mean C value yields $F(4,55) = 3.75$ which is significant beyond the .01 level. Tukey tests using the 95% confidence interval on all possible pairs of comparisons among groups show only the difference between the W± and W+ group to be significant. Thus, the analysis on the C parameter measure bears out the difference between the two reference groups, but does not provide the basis for a decision concerning the three experimental groups.

The experimental groups were evaluated with respect to the W+ and W± groups by two further analyses using the $4CA - B^2$ measure. This measure, however, is sensitive to the value of C used. So Table 1 indicates two sets of means and *SD*s for $(4CA - B^2)$ using, first, Mean $C = 0.608$ computed from Groups W+, w+, WB, and WG and second, Mean $C = 0.330$ computed from Groups W±, w+, WB, and WG. Simple analysis of variance on the first measure yields nonsignificant differences among groups, $F(3,44) = 1.10$, $p < .20$. For the second measure, significant differences among groups are obtained, $F(3,44) = 7.33$, $p < .001$, and subsequent Tukey tests using the 95% confidence interval show all three experimental groups to differ significantly from W±, but not to differ among themselves.

The results of the additional analyses performed to control for different terminal acquisition levels among groups correspond to the results of the original analysis; namely, Group W± differs from Group W+, and Groups w+, WB, and WG are more similar to Group W+ than to Group W±. Experiment II, employing only Groups WB and WG, was carried out to obtain further data on the terminal acquisition response speed and its relation to extinction for these two conditions.

EXPERIMENT II

A suggested explanation[2] for the slower terminal acquisition speeds of the discrimination groups (particularly WG) is that *S*s in these groups are exposed to two stimuli over the course of training while *S*s in the other groups are exposed to only one. Since the latency of response is measured from the time the goal platform stimulus is presented until *S* leaves the approach platform, *S*s may respond more slowly in the discrimination groups because time is taken for "stimulus processing" (e.g. in the making of observing responses). If this were the case, it might also be expected that more time would be taken by the WG than the WB group, a conjecture that is compatible with the findings of Exp. I.

---

[2] G. S. Blum and W. J. McKeachie.

In Exp. II, a larger goal platform is used and both the running speed while on the goal platform and response speed in leaving the approach platform are recorded. The latter measure is used as a replication for Exp. I, and the new measure is intended to minimize the "stimulus processing" time referred to since it is taken after *S* enters upon the goal platform and is presumably responding to the cues of that platform. The acquisition and extinction performance of the two groups, WG and WB, are compared using both measures.

Method

Subjects.—The *S*s were 27 experimentally naive male pigmented rats from the colony maintained by the Michigan Psychology Department. The *S*s were approximately 90 days old at the beginning of training. Two *S*s were discarded, 1 from Group WB for refusal to run and 1 from Group WG because of apparatus failure, leaving the groups with 12 and 13 *S*s, respectively.

Apparatus.—The elevated runway of Exp. I was adapted for Exp. II by the addition of two photoelectric cell units which are mounted externally to the apparatus housing and which control stopping the two Standard Electric clocks. The sequence of *S*'s behavior and the resulting time measures are as follows: *S* first breaks the infrared beam located at the end of the approach platform starting three clocks and turning on the goal chamber illumination. The first clock stops by a microswitch arrangement when *S* leaves the approach platform, the second clock stops when the infrared beam located 8 in. from the beginning of the goal platform is broken, and the third clock stops when *S* interrupts the infrared beam 7 in. from the end of the goal platform. The time measurements of interest are the latency of *S* in leaving the approach platform (the same measure as in Exp. I) and the time taken by *S* to travel the 10-in. section of the goal platform. The latter measure is obtained by subtracting the second clock reading from the third clock reading. The goal platform is 25 in. long and 4 in. wide and painted either black, gray, or white as in Exp. I.

Procedure.—The procedures of Exp. I were carried out in Exp. II with one exception. In Exp. II both groups were given 26 acquisition trials in the sequence, + − + + − − + − − − + + − + − + − − + + + − + + − −, on Day 1. On Day 2, 8 additional acquisition trials in the sequence, + − + + − − + −, were provided prior to 19 extinction trials.

Results

Figure 2 shows both time measures in terms of mean log response speed during extinction for the two groups. The same kinds of analyses were carried out on these data as in Exp. I. Table 2 summarizes the data. The results are as follows:

1. Overall differences between the speed to the positive and negative

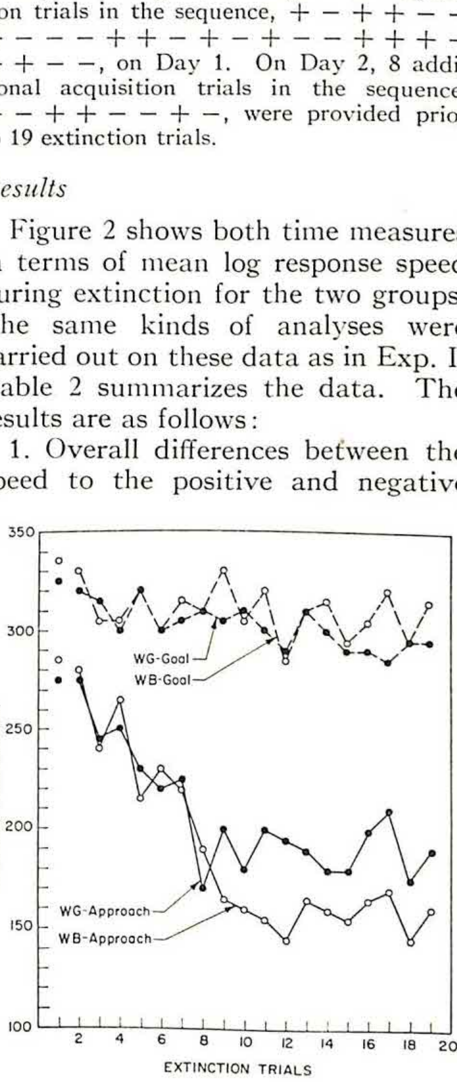


FIG. 2. Extinction performance of the WB and WG discrimination groups with the approach and the goal measures.

TABLE 2

MEANS AND *SD*s OF MEASURES APPROPRIATE TO EVALUATION OF EXTINCTION PERFORMANCE FOR APPROACH AND GOAL PLATFORM SPEEDS

Measure		Approach Speed		Goal Platform Speed	
		Group WG	Group WB	Group WG	Group WB
Extinction speed	Mean *SD*	206 38.0	188 60.3	303 26.5	310 17.7
Terminal acquisition speed	Mean *SD*	281 44.8	270 68.6	322 13.8	333 10.4
C	Mean *SD*	.58 .64	.71 .42	.02 .50	.04 .32
B	Mean *SD*	−4.08 5.00	−6.71 4.07	−1.51 2.04	−.54 1.76
A	Mean *SD*	190 44.8	168 64.1	302 35.1	308 22.7
$(2.764)A - B^2$	Mean *SD*	484 99	404 184		
$(.126)A - B^2$	Mean *SD*			32 11	36 6

stimuli favor the positive stimulus in both groups on the approach platform measure, $t(11) = 2.63$, $p < .05$ for WB and $t(12) = 1.41$, $p < .10$ for WG, but do not agree in direction with the goal platform measure, $t(11) = 1.17$, $p > .20$ for WB and $t(12) = -0.74$, $p > .40$ for WG.

2. The mean difference in speed to the positive and negative stimuli is greater for WB than for WG, but the means for the two groups do not differ significantly, $F(1,23) < 1.00$ for the approach measure and $F(1,23) = 1.95$, $p > .10$, for the goal platform measure.

3. Differences between terminal acquisition speeds, to the positive stimulus for the two groups, as suggested in Exp. I, were not found for the approach measure, $F(1,23) < 1.00$, but favor the WB group with the goal platform measure, $F(1,23) = 4.43$, $p < .05$. No differences between the two groups during extinction approach significance with either platform measure by analysis of log response speed, the C values, or $4CA - B^2$.

DISCUSSION

The performance of the groups during extinction shows (*a*) that the WG and WB groups do not differ from each other and (*b*) that both are more similar to the W+ and w+ groups than they are to the W ± group. Thus, for the conditions of the present experiments, no resistance to extinction beyond that appropriate to a 100% reinforced stimulus can be attributed to the positive stimulus of the discriminations.

Jenkins (1961), however, using the key press response of pigeons, has demonstrated greater resistance to extinction for the positive stimulus of a discrimination than for the same stimulus under simple 100% reinforcement. In addition to the apparent difference in *S*s and apparatus, three procedural differences between the studies may be particularly important.

First, the difficulty of the discrimination was different in the two cases. Jenkins used stimuli which were small horizontal or vertical dark lines on a back lighted key. That a relatively difficult discrimination problem resulted is seen from his report that 17–25 training sessions, each with a minimum of 40 trials, were required to reach criterion. This contrasts with the present experiments in which clear evidence of discrimination appeared after only 36 trials.

Second, Jenkins provided 5–7 days of 100% reinforced training to the positive stimulus alone prior to discrimination learning. Such pretraining was not given in the present study.

A third difference in the experiments is in the procedure during extinction. The only change Jenkins made in going from acquisition to extinction sessions was to remove the reward. The discrimination group continued to be presented with the negative stimulus just as during acquisition. This produced different extinction conditions for his three groups (intermittent, 100%, and discrimination). In the present experiments, only the positive stimulus was presented during extinction.

The failure to replicate in Exp. II the difference in the speed of approach to the white platform at the end of acquisition for Groups WB and WG, suggested in Exp. I, warrants brief mention. Examination of all four groups points to WG in Exp. I as the deviant group. Differences in *S* population, acquisition procedures, and length of goal platform were present in the two experiments but suggest no compelling hypothesis to account for the result. The conclusions concerning the resistance to extinction of the WG and WB groups are the same despite the differences in acquisition results in the two experiments.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

Anderson, R. L., & Houseman, E. E. Tables of orthogonal polynomial values extended to $N = 104$. *Ia. Agric. Exp. Sta., res. Bull.*, 1942, **297**, 595–672.

Birch, D., Ison, J. R., & Sperling, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

Jenkins, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, **61**, 111–121.

Scheffé, H. *The analysis of variance.* New York: Wiley, 1959.

Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century-Crofts, 1938.

Wickens, D. D., & Snide, J. D. The influence of nonreinforcement of a component of a complex stimulus on resistance to extinction of the complex itself. *J. exp. Psychol.*, 1955, **49**, 257–259.

(Received February 16, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 156–162

PRIOR EXPERIENCE AS A DETERMINANT OF FIGURE-GROUND ORGANIZATION

HENRY G. CORNWELL

Lincoln University

The failure to reproduce Rubin's aftereffect reported by Rock and Kremen (1957) was hypothesized as due to the absence from their training figures of the same black-white context that existed in their ambiguous test figures. This context was provided in the present experiment by training Ss in figure perception with reversible but unbalanced black-white designs, and subsequently testing them with identical contours in balanced designs. Comparison (chi square) of experimental and control Ss' responses to test designs revealed a significant effect ($p < .02$): Ss perceived the same figure-ground organization in the ambiguous test designs that they had perceived in the training designs.

Recently doubt has been expressed concerning the validity of the conclusions reached by Rubin (1921) as to the effect of prior experience on figure-ground organization. In Rubin's experiment *S*s were instructed to perceive a given part of an ambiguous design as figure. After extensive training the *S*s were later shown the same designs interspersed with other ambiguous designs that they had not previously seen, and were asked to report in what part of each design they perceived figure. Upon the basis of his results Rubin concluded that the prior training with the designs produced a significant effect on the way in which they were subsequently perceived. Similar conclusions have been reached by Schafer and Murphy (1943) concerning the effect of prior reward, and by Smith and Hochberg (1954) concerning the effect of prior punishment on the subsequent perception of rewarded or nonpunished halves of ambiguous and reversible facial profiles. Rock and Kremen (1957) carefully reviewed Rubin's experiment, and concluded that several defects in his procedure made it difficult to accept his results as convincing evidence of the effect he reported. Modifying Rubin's procedure, these investigators trained *S*s on one of the halves of 8 ambiguous designs, and after a 5-min. intermission they presented test designs containing both halves of the reversible figures interspersed with 10 control designs that had not been seen during training. Negative results were obtained.

The present investigator conceives that the crucial condition for producing such an effect as that reported by Rubin is the preservation of the same contextual relationship from the unambiguous training situation to the ambiguous test situation. This condition was satisfied in each of the similar investigations reporting such an effect. Rock and Kremen, however, in making certain that their *S*s could see only the desired halves as figure during training, eliminated from the training situation the black-white contextual relationship which is here hypothesized as necessary for the production of the effect. The present experiment, therefore, is a replication of the Rock and Kremen experiment, in which the training procedure has been modified so as to provide the same contextual relation-

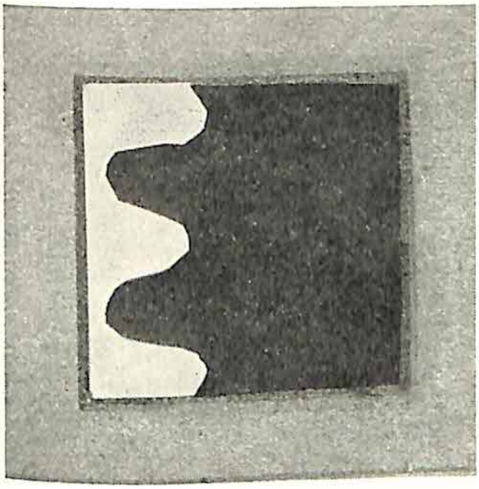


FIG. 1. One of the eight unbalanced training designs.

ship during training that will exist in the test, and at the same time afford a partial control over what the *S*s perceive during training. It is assumed that both of these objectives can be achieved by presenting spatially unbalanced contours in the training situation, and the same contours spatially balanced in the test. Specifically, the hypothesis proposed is that if a contour between black and white areas is presented tachistoscopically and the black and white areas are strikingly unbalanced or unequal, a given *S* will tend to be consistent in perceiving one of the areas as figure. If the same contour is later presented between black and white areas that are approximately balanced or equal, the *S* will tend to perceive the same area as figure that he perceived when the two areas were unequal.

METHOD

Materials.—Facsimiles of the eight test figures published by Rock and Kremen (1957) were used in the present experiment. Stimulus targets were constructed of 3-in. squares of neutral gray cardboard, on which were mounted 2-in. squares containing black and white areas separated by a contour. Eight training targets were constructed so that either the black or white area was greatly in excess of the other. The white area was greater than the black in half of the targets, and the black was greater than the white in the other half. Both white and black appeared equally often on the right and left, and each combination of size and color of area appeared equally often in both positions. One of the training targets is shown in Fig. 1.

Eighteen test targets were constructed the same size as the training cards. Eight of the test targets contained the same contours as the training designs, but the central contour separated equal areas of black and white in each case. The remaining 10 test targets were designated control designs, as they contained contours unlike any of those occurring in the training designs. The black and white areas in the control designs were also equal in each case. Examples of test and control designs are shown in Fig. 2 and 3. A demonstration target unlike any of the training or test designs was also prepared. Photometric determination of the percent reflectances of the white, gray, and black components of the targets showed these to be, respectively, 84%, 39%, and 7%. Targets were presented at a distance of 38 in. from the *S*'s eyes by means of a falling screen tachistoscope adjusted for 1-sec. exposures. All observations of the training and test designs were binocular.

Subjects.—The *S*s were 40 Negro male college students from all classes at Lincoln University. They were randomly assigned to compose an experimental and a control group of 20 *S*s each. Experimental and

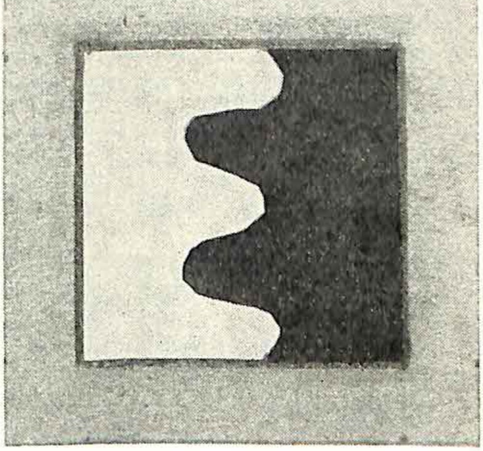


FIG. 2. Ambiguous test design corresponding to the training design.

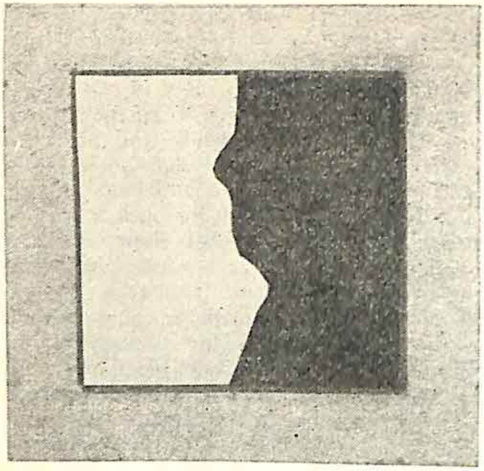


FIG. 3. One of the 10 ambiguous control designs.

control *S*s were alternately scheduled for sessions in the laboratory. The *S*s were paid for their services.

Procedure.—Each *S* was shown a reproduction of Rubin's reversible vase-profiles figure and given a demonstration of how either the black or white portion could be seen as figure. Then he was read the following instructions:

> This is an experiment to see how certain factors affect your perception of figures. You will be shown designs like this one very rapidly, and you are to tell me which part of the design, the black or the white, looks like figure rather than like the background. A figure usually looks as though it is in front of the background. Also a figure seems to have a definite shape while the background does not. As soon as you see a design you are to say out loud whether the white or the black part looks like the figure. If both parts, black and white, look like figures, you may say, "Both." If neither the white nor the black looks like a figure, you may say, "Neither." Now look at this card and tell me which part looks like a figure.

The *S* was seated comfortably in front of the tachistoscope with his chin resting on his folded arms, which in turn rested on the end of the laboratory table. By looking straight ahead at eye level the *S* saw a 3-in. square of neutral gray cardboard against a black screen under a constant incandescent illuminance of 1.2 ft-c. This provided a pre-exposure field. After a ready signal the lower screen was released to expose the demonstration target, which was the same size and in the same position as the gray pre-exposure square. The demonstration target was visible for 1 sec. and was then covered by the upper screen, upon which a 3-in. gray cardboard square provided a postexposure field. Each experimental *S* was then shown the eight unbalanced training designs eight times over in a counterbalanced order, so that each design appeared once in each of eight serial positions. The *S*'s response to each design was recorded after its number on a record form as B (black), W (white), = (both), and O (neither). At the end of the training series the *S* was told:

"Now we will do something a little different. I shall show you some sentences in the same way to see how well you can read them."

A 15-min. interval was filled with the repeated tachistoscopic presentation of from one to five short sentences typed on white cards until each was read aloud without error. The *S* was then told:

"Now we will see some more designs. Again you are to tell me which part of the design looks like a figure, just as you did before."

The 8 test designs containing the same contours as the training designs were presented interspersed with the 10 control designs in such a way that no 2 test designs occurred in succession. The first two targets in the test series were always control designs. The 18 designs were then randomly spread out on the laboratory table and *S* was instructed to pick out and hand to *E* any that he recognized as having been presented before the sentences were shown, and to specify the part of the design that seemed familiar to him. Designs and parts recognized were checked on the record sheet.

Control *S*s were given the same initial instructions and shown the demonstration target, after which they were told:

"Before we do more like this, I shall show you some sentences in the same way to see how well you can read them."

A 15-min. interval of sentence presentation was followed by presentation of the 18 test and control designs, which concluded the session for control *S*s. All *S*s were requested not to discuss the experiment.

RESULTS

Training data.—In order to determine whether any of the eight unbalanced training designs were perceived in the same way by all *S*s, a tabulation was made of the frequency with which each *S* saw the black, white, neither, or both areas of the training designs as figure. Inspection of Table 1 shows that no training design was consistently perceived by a significant majority of *S*s in the same way. After eliminating *S*s who saw both parts of a design equally often as figure, comparison of the number who saw the black area of a given design as figure with the number who saw the white area of the same design as figure shows no significant departure of the obtained frequencies from frequencies expected upon the basis of an equal probability hypothesis. The design showing the greatest variation from the expectancy 9B–9W is Training Design 1 with Frequencies 6B–12W, $\chi^2(1) = 2.00$, $p < .20$.

In order to evaluate the effect of the prior training upon the subsequent perception of figure in the ambiguous test designs it was necessary to identify and classify together all *S*s perceiving each training design in the same way. Accordingly, a tabulation by training design of the response frequencies of all *S*s perceiving the black area as figure more often than the white was made; a similar tabulation for *S*s perceiving the white area as figure more often than the black was made. The total responses to each training design for these two categories are shown in Table 2. "Both" and "Neither" responses have been omitted from this and all subsequent tabulations and evaluations.

TABLE 1

NUMBER OF *S*S SEEING BLACK OR WHITE AREA OF THE EIGHT TRAINING DESIGNS AS FIGURE

Training Design	*Ss* Seeing Black as Figure	*Ss* Seeing White as Figure	*Ss* Seeing Both or Neither as Figure
1	6	12	2
2	7	11	2
3	11	8	1
4	8	11	1
5	10	8	2
6	9	9	2
7	11	9	0
8	12	8	0

TABLE 2

TRAINING DATA IN RESPONSE FREQUENCIES

Training Design	*Ss* Seeing Black as Figure		*Ss* Seeing White as Figure	
	Black	White	Black	White
1	39	8	11	75
2	42	8	10	71
3	75	6	7	42
4	46	9	11	65
5	54	5	11	36
6	51	11	4	60
7	63	17	3	61
8	74	12	4	47
Total	444	76	61	457

Ambiguous test design data.—Evaluation of the effect of prior training requires the comparison of responses made by the experimental *S*s to the ambiguous test designs with responses made to the same designs by control *S*s who received no prior training. Therefore, tabulations were made of responses to the same test designs for experimental *S*s and for control *S*s corresponding to the experimental *S*s in terms of order of treatment in the laboratory. For example, the responses to Test Design 1 made by Experimental *S*s 3, 4, 6, 13, 16, and 17, who saw the black area of Training Design 1 as figure, were compared with the responses to Test Design 1 made by Control *S*s 3, 4, 6, 13, 16, and 17.

Table 3 shows the responses made by experimental and control *S*s to the 8 ambiguous test designs and to the 10 control designs. Comparison of the test design totals for the experimental *S*s who saw the black area of the training designs as figure (distribution in upper half of first column; Totals

TABLE 3

Frequency of Figure Perception in Black and White Areas of the Ambiguous Designs by Experimental and Control *S*s

	Experimental *S*s (Training B)		Corresponding Control *S*s		Experimental *S*s (Training W)		Corresponding Control *S*s	
	Black	White	Black	White	Black	White	Black	White
				Test Designs				
1	6	0	2	3	2	8	2	9
2	5	1	4	2	4	6	5	6
3	10	0	7	4	0	5	4	3
4	4	1	3	4	1	9	2	7
5	4	0	5	4	1	4	5	1
6	7	0	4	5	2	6	3	5
7	7	3	3	7	0	7	1	7
8	8	2	7	3	1	2	4	3
Total	51	7	35	32	11	47	26	41
				Control Designs				
1	5	5	7	5	4	5	7	3
2	9	3	10	3	7	3	10	2
3	7	1	4	2	6	1	3	1
4	7	2	8	4	6	2	7	4
5	4	6	3	4	3	6	2	2
6	6	8	4	8	5	8	3	8
7	6	2	6	7	5	1	6	6
8	7	5	7	7	5	6	6	8
9	1	1	1	8	0	1	1	6
10	3	3	4	4	0	3	4	3
Total	55	36	54	52	41	36	49	43

51B–7W) with the same test design totals for corresponding control *S*s (distribution in upper half of second column; Totals 35B–32W) reveals a significant effect, $\chi^2(1) = 18.46$, $p < .001$. Similarly, comparison of the test design totals for experimental *S*s who saw the white area of the training designs as figure (distribution in upper half of third column; Totals 11B–47W) with the test design totals for corresponding control *S*s (distribution in upper half of fourth column; Totals 26B–41W) shows a significant effect, $\chi^2(1) = 5.93$, $p < .05$. Comparisons of responses made by experimental *S*s to control designs with responses made by corresponding control *S*s to control designs (distribution in lower half of first column having Totals 55B–36W with distribution in lower half of second column having Totals 54B–52W; and distribution in lower half of third column having Totals 41B–36W with distribution in lower half of fourth column having Totals 49B–43W) show no significant differences, $\chi^2(1) = 1.80$; $\chi^2(1) = 0.00$. These comparisons and evaluations are summarized in Table 4.

Recognition data.—Table 5 shows the frequency with which each test design was "recognized" as having been seen previously during training, and identifies the part of the design recognized as to how it was perceived on the test trial. Entries in the table for control designs are false recognitions. Analysis of the totals shows that in 96 out of 160 test design presentations correct recognitions were obtained, 65 of these being in the area perceived as figure and 31 in the area perceived as ground. In 21 out of 200 control design presentations false recognitions were obtained. The difference between test design recognition and control design false recognition totals reveals a significant recognition effect, $\chi^2(1) = 99.13$, $p < .001$, for the test designs which may be attributed only to the prior experience with the corresponding training designs.

TABLE 5

RECOGNITION DATA

	Part of Design Recognized		Total Recognitions
	Figure	Ground	
	Test Designs		
1	5	4	9
2	11	1	12
3	11	6	17
4	7	3	10
5	7	7	14
6	10	1	11
7	11	2	13
8	3	7	10
Total	65	31	96
	Control Designs		
1	0	1	1
2	2	0	2
3	1	2	3
4	0	3	3
5	0	0	0
6	1	2	3
7	1	2	3
8	1	3	4
9	0	0	0
10	1	1	2
Total	7	14	21

DISCUSSION

The results clearly show a significant effect produced by the prior training with the unbalanced designs upon the subsequent perception of figure in the ambiguous balanced designs. This effect was observed when experimental *S*s' responses to test designs were compared with corresponding control *S*s' responses to the same designs. It appears, then, that one of the necessary conditions for the demonstration of the effect originally reported by Rubin (1921) is the preservation of the same context in the stimulus designs from the prior training to the subsequent test.

TABLE 4

SUMMARY OF COMPARISONS BETWEEN EXPERIMENTAL AND CONTROL RESPONSES

Comparison	χ^2
Experimental (B) *S*s' test designs with corresponding control *S*s' test designs	18.46***
Experimental (W) *S*s' test designs with corresponding control *S*s' test designs	5.93*
Experimental (B) *S*s' control designs with corresponding control *S*s' control designs	1.80
Experimental (W) *S*s' control designs with corresponding control *S*s' control designs	0.00

* $p < .05$.
*** $p < .001$.

A plausible explanation of the results obtained in this experiment, and in the earlier studies cited, would postulate the development of a specific set or perceptual bias for each training design. In terms of trace theory this specific set could be correlated with the development of the trace of a relationship in which a contour functions in a given way. The recognition data confirm the effectiveness of the training experience in establishing cues for the recognition of the designs. Although the *S*s reported recognition of figure twice as often as recognition of ground, the fact that a third of the recognitions appeared in the ground suggests strongly that neither the figure nor the ground organization is the crucial determinant of recognition, but that the contour is the element that is recognized. This interpretation answers the question raised by Rock and Kremen (1957) as to how any relevant trace can enter into the process that determines a percept if that trace cannot be aroused until after the form (figure) is perceived. It is suggested that the contour, not the figure, is first perceived and recognized, thereby reactivating a relational trace, which in turn determines which region of the ambiguous stimulus field will be perceived as figure.

REFERENCES

ROCK, I., & KREMEN, I. A re-examination of Rubin's figural aftereffect. *J. exp. Psychol.*, 1957, **53**, 23–30.

RUBIN, E. Visuell wahrgenommene Figuren. Copenhagen: Gyldendalske, 1921.

SCHAFER, R., & MURPHY, G. The role of autism in a visual figure-ground relationship. *J. exp. Psychol.*, 1943, **32**, 335–343.

SMITH, D. E. P., & HOCHBERG, J. E. The effect of "punishment" (electric shock) on figure-ground perception. *J. Psychol.*, 1954, **38**, 83–87.

(Received February 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 163–168

AN ANALYSIS OF SOME FACTORS UNDERLYING INTRALIST ASSOCIATIVE TRANSFER IN PAIRED-ASSOCIATE LEARNING [1]

JOHN C. McCULLERS [2]

University of Texas

A within-list method proposed by Spence for varying associative strength and degree of competition in paired-associate learning served as the basis of a logical analysis. With Spence's method as the basic model, the analysis resulted in the formulation of 2 new models. 48 6th grade children served as Ss in an investigation of these 3 models. The results supported an hypothesis that number and type of competitive connections are important in determining the degree of intralist associative transfer. A general implication was that the new models might prove less complicated than the original method for the purpose of assessing hypotheses relating to the interaction of habit and drive.

This study is concerned with the role of selected factors as determinants of intralist associative transfer in verbal paired-associate learning. It is based on extensions of a within-list method originally proposed by Spence (1956) for varying the factors of associative strength and interference.

Basically, Spence's method consists in constructing the paired-associate list in such a way that the stimulus and response items of certain word pairs are linked by initially strong associative connections while the remaining pairs have initially weak connections. Further, the stimulus items of the strong and weak pairs are synonymous, so that a strong associative connection also exists between the stimulus item of a weak word pair and the response item of a strong word pair. A diagram and example of this method are given below:

$$\begin{tikzcd} S_A \arrow[r] & R_A & \text{BEAUTIFUL} \arrow[r] & \text{UGLY} \\ S_{A'} \arrow[r, dashed] \arrow[ur] & R_B & \text{PRETTY} \arrow[r, dashed] \arrow[ur] & \text{SHIRT} \end{tikzcd}$$

The solid-shaft arrows indicate initially strong associative connections and the broken-shaft arrows indicate initially weak connections. According to theory, learning of the weak S-R pair would be expected to occur in the presence of strong competition produced by $S_{A'} \rightarrow R_A$. Early in training, learning of the strong pair ($S_A \rightarrow R_A$) would be expected to proceed with minimal interference due to the initially strong associative connection and the absence of any initially strong competing connection. However, an hypothesis has been advanced (Ramond, 1953; Spence, Farber, & McFann, 1956) which holds that as the weak pair is learned, a generalized association may develop between S_A and R_B which would impair performance on the strong pair later in training.

The purpose of the present paper is to illustrate that within Spence's original method there are logically two sources of competition present. For example, in addition to the strong connections indicated, i.e., $S_A \rightarrow R_A$ (BEAUTIFUL → UGLY) and $S_{A'} \rightarrow R_A$ (PRETTY → UGLY), an initially strong connection would also seem to exist, though not shown, between the two synonymous stimulus items, i.e., $S_A \rightarrow S_{A'}$ (BEAUTIFUL → PRETTY). Thus, one source of competition arises from the presence of an interstimulus item association and another arises from an inter-word-pair S-R association. Since two words may be strongly associated for reasons other than synonymity; e.g., homonymity, functional relationships, etc., it would appear possible to construct a method having only one or the other of these two sources of competition present. With Spence's method, redrawn to show the interstimulus item connection, as the basic model (I), one might construct a model (II) in which the stimuli are strongly associated but in which there is no strong connection between $S_{A'}$ and R_A. Similarly, one might construct a model (III) in which both stimuli are strongly associated with R_A but are not in themselves associated. Diagrams of these three models are presented in Fig. 1.

---

[1] This study is based on a PhD dissertation submitted to the Graduate School of the University of Texas. The research was supported in part by a United States Public Health Service Predoctoral Research Fellowship (MF-13000) from the National Institute of Mental Health and in part by an NIMH grant (M-4240) to Alfred Castaneda, to whom the author is indebted for supervision. The manuscript was written at the Institute of Child Development, University of Minnesota, during the tenure of a Postdoctoral Research Fellowship (MF-13000-C1) from NIMH.

[2] Now at San Jose State College.

A word is perhaps in order concerning the reversibility of associations. Although the arrows employed in the diagrams imply a unidirectionality of association, it is obvious that this is not always the case. Since the nature of the task ensures that the stimuli always precede the response items, the problem of directionality is perhaps not as critical in the case of S-R associations. With respect to the strongly associated, synonymous stimuli, however, it is apparent that the association $S_{A'} \rightarrow S_A$ (PRETTY → BEAUTIFUL) is no less important as a source of interference than $S_A \rightarrow S_{A'}$ (BEAUTIFUL → PRETTY). Accordingly, $S_A \rightarrow S_{A'}$ should be considered as a source of competition to the learning of the strong word pair and $S_{A'} \rightarrow S_A$ as a source of competition to the learning of the weak pair.

With respect to learning the strong pair, the interstimulus item connections of Models I and II would be expected to serve as: (*a*) direct sources of interference which would produce stimulus errors and (*b*) indirect sources of interference which might provide for generalized response errors later in training as the weak-pair associative connection becomes stronger. With respect to learning the weak pair, the interstimulus item connections of Models I and II would be expected to serve as direct sources of stimulus errors and indirect sources of generalized response errors. The inter-word-pair S-R connections of Models I and III would be expected to serve as direct sources of response errors.

It might be predicted, therefore, that *S*s would make: (*a*) more stimu-

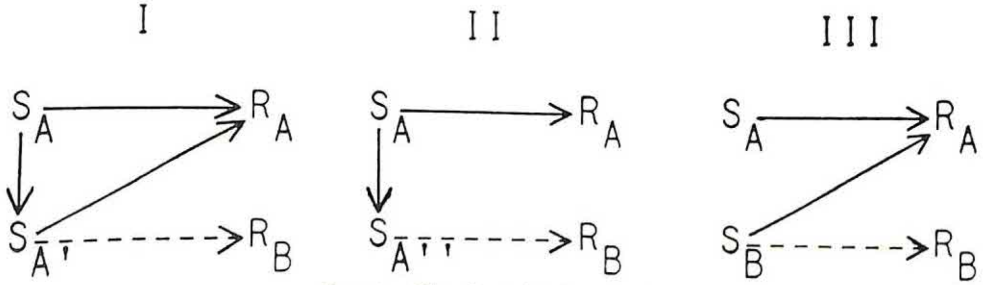


FIG. 1. The three basic models.

lus errors on both strong and weak word pairs in Models I and II than in III, (*b*) more generalized response errors on the strong pairs in Models I and II than in III, and (*c*) more response errors on the weak pairs in Models I and III than in II. Further, since the generalized support (through $S_{A'} \rightarrow S_A$ and $S_A \rightarrow R_A$) of the competitive connection $S_{A'} \rightarrow R_A$ in Model I would tend to maintain $S_{A'} \rightarrow R_A$ as a strong source of competition throughout training, *S*s would be expected to make more response errors on the weak pairs in Model I than in Model III.

METHOD

Subjects.—The *S*s consisted of 48 sixth grade children[3] possessing a minimum of fourth grade reading ability. Sixteen *S*s, 8 males and 8 females, were randomly assigned to each of the three basic conditions.

Apparatus and material.—A Hunter Card Master (Model 340) was used to present the task lists. The task words were selected from a list standardized on fourth, fifth, and sixth grade children (Castaneda, Fahel, & Odom, 1961). The S-R pairs were constructed in accord with the diagrams of Fig. 1 and are presented in Table 1. Four sets of word pairs were employed in each condition with each set consisting of a strong S-R pair and a weak pair immediately below. The associative strengths of the strong S-R pairs and various competitive connections were approximately 20%; weak pairs carried initial associative strengths of approximately 1%. The association between the words of different sets in each condition was initially low (approximately 1%).

Procedure.—Each *S* participated individually and was instructed as follows:

> This is a game to see how well you can learn. This is how it goes. Here on this side (*E* points to left aperture) a little window will open up and you will see a word. A few seconds after this window opens, the window on this side (*E* points to right aperture) will open and you will see a *second* word. Your job is to guess what the second word is before its window opens. You say the word out loud so that I can hear you. Do you think you understand how it goes now? Fine; now we are ready to start. Remember, you are to say out loud the second word which is going to appear in the second window *before* the window opens. In the beginning you will make a lot of mistakes, but as you learn how it goes you won't make so many mistakes. Ready? Any Questions?

TABLE 1

WORD PAIRS FOR EACH CONDITION

Stimulus			Response
Cond. I	Cond. II	Cond. III	
LONG[a]	LONG[a]	LONG[a]	SHORT
TALL	BIG	LITTLE	BROKEN
HOT[a]	HOT[a]	HOT[a]	COLD
WARM	BOILING	FREEZING	FUNNY
LIGHT[a]	LIGHT[a]	LIGHT[a]	DARK
BRIGHT	SUN	BLACK	REAL
PRETTY[a]	PRETTY[a]	PRETTY[a]	NICE
BEAUTIFUL	UGLY	GOOD	SAME

[a] Strong associative pairs.

Each stimulus item was exposed for a 4-sec. period consisting of a 2-sec. anticipation interval and a 2-sec. simultaneous presentation with the response item. A 2-sec. interpair interval in which both windows of the apparatus were closed followed the exposure of each word pair. The pairs of each list were presented randomly and, to minimize serial learning, three different orders of exposing the pairs were employed. Each *S* was taken to a criterion of 1 errorless trial or to a maximum of 24 trials. A trial consisted of one exposure of each of the eight word pairs. Verbatim responses were recorded on each trial. Those *S*s who failed to anticipate correctly one or more items during the 24 trials were given an additional presentation and asked to read the words aloud. This was done to determine whether the *S*'s failure to reach criterion was due to an inability to read the words. This procedure was necessary in the case of 3 *S*s, all girls, each of whom read the words satisfactorily.

RESULTS

Correct anticipations.—A four-factor mixed-design analysis of variance was performed using number of

---

[3] Special thanks are extended to C. E. Ellison, Superintendent, Killeen Public Schools; to Marcus J. Voltin, Principal, East Ward Elementary School, Killeen, Texas; and to Leila S. Fahel of the Office of Child Research, the University of Texas, for their assistance in making *S*s available.

correct anticipations as scores. Sex and Conditions were analyzed as "between" factors; Associative Strength of S-R pair and Trials in blocks of four were analyzed as "within" factors. A summary of this analysis is presented in Table 2.

Mean percent correct anticipations were computed separately for each type of S-R pair within each condition. This information plotted across blocks of trials is presented in Fig. 2.

Orthogonal comparisons, each based on 1 and 570 *df*'s, revealed that performance on the weak pairs in Cond. II and III was significantly better than in Cond. I ($F = 27.15$). Also, performance on the weak pairs in Cond. II was better than in Cond. III ($F = 15.45$). Performance on the

TABLE 2

Analysis of Variance of Correct Anticipations

Source	*df*	*MS*	*F*
Between *S*s	47		
Sex (S)	1	112.01	2.63
Conditions (C)	2	226.90	5.32**
S × C	2	21.00	<1.00
Error (b)	42	42.66	
Within *S*s	528		
Associative strength (A)	1	3364.00	280.10***
Trials (T)	5	1008.54	271.11***
A × T	5	180.00	56.25***
S × A	1	24.17	2.01
S × T	5	3.95	1.06
C × A	2	156.82	13.06***
C × T	10	6.68	1.80
S × C × A	2	11.20	<1.00
S × C × T	10	5.97	1.60
S × A × T	5	7.69	2.40*
C × A × T	10	5.42	1.69
S × C × A × T	10	2.99	<1.00
Error 1 (w)	42	12.01	
Error 2 (w)	210	3.72	
Error 3 (w)	210	3.20	
Total	575		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 3

Analysis of Variance of Errors

Source	*df*	*MS*	*F*
Between *S*s	47		
Conditions (C)	2	56.26	5.14**
Error (b)	45	10.95	
Within *S*s	2256		
Associative strength (A)	1	827.76	262.78***
Trials (T)	5	254.25	267.63***
Type of error (E)	3	223.92	30.34***
A × T	5	45.38	56.02***
A × E	3	70.52	16.91***
T × E	15	23.22	9.84***
A × T × E	15	4.69	6.90***
C × A	2	38.80	12.32***
C × T	10	1.74	1.83
C × E	6	43.40	5.88***
C × A × T	10	1.34	1.65
C × A × E	6	30.12	7.22***
C × T × E	30	3.64	1.54*
C × A × T × E	30	2.78	4.09***
Error 1 (w)	45	3.15	
Error 2 (w)	225	0.95	
Error 3 (w)	135	7.38	
Error 4 (w)	225	0.81	
Error 5 (w)	135	4.17	
Error 6 (w)	675	2.36	
Error 7 (w)	675	0.68	
Total	2303		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

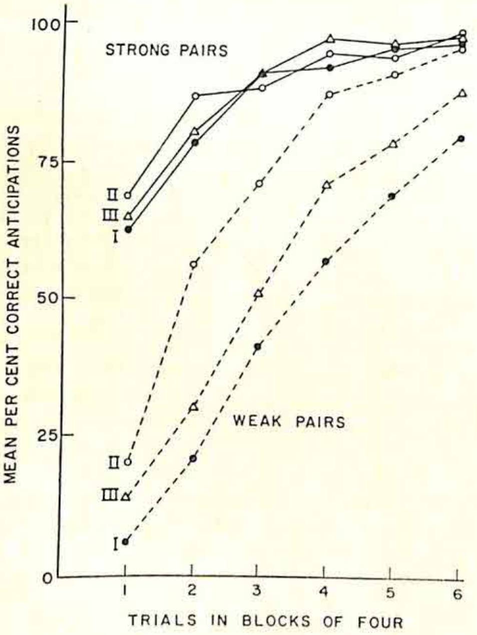


Fig. 2. Mean percent correct anticipations plotted separately across trials for the strong and weak pairs of Cond. I, II, and III.

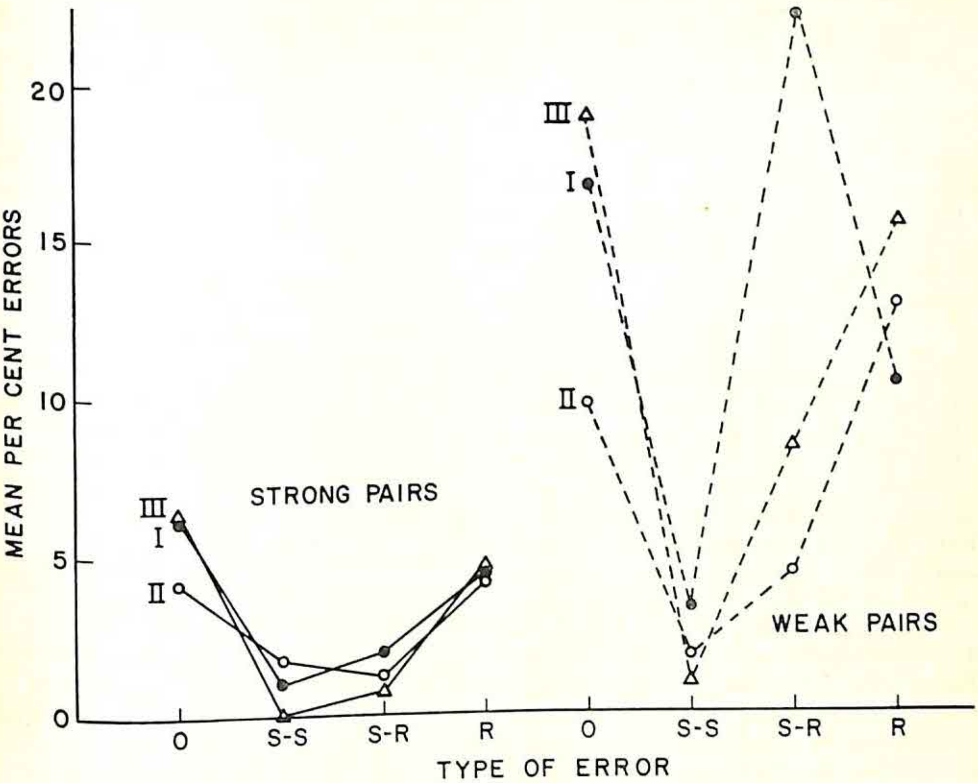


FIG. 3. Mean percent errors of omission (O), stimulus errors (S-S), response errors (S-R), and random errors (R) plotted separately for the strong and weak pairs of Cond. I, II, and III.

strong pairs did not differ significantly among the three conditions ($F < 1.00$).

Error scores.—A second analysis of variance was performed on error scores in which four types of errors were analyzed. These were: (*a*) errors of omission, in which the stimulus item elicited no overt response during the anticipation interval, (*b*) stimulus errors, in which one stimulus item elicited the other stimulus item in its set, (*c*) response errors, in which a stimulus item elicited the response item of the other pair in its set, and (*d*) random errors, in which the stimulus item elicited either a word of another set or some nontask word. A summary of this analysis is presented in Table 3.

Orthogonal comparisons, each based on 1 and 2160 *df*'s, were performed across blocks of four trials among the strong and weak pairs for the three conditions. With respect to the strong pairs, significantly more stimulus errors occurred in Cond. I and II than in III at the first block of trials ($F = 5.61$). With respect to the weak pairs, more stimulus errors occurred in Cond. I and II than in III at the first block of trials ($F = 5.93$). More response errors occurred in Cond. I and III than in II at every block of trials except the last: Block 1 ($F = 19.23$), Block 2 ($F = 42.81$), Block 3 ($F = 16.87$), Block 4 ($F = 23.07$), and Block 5 ($F = 8.14$). Further, more response errors occurred in Cond. I than in III at every block of trials: Block 1 ($F = 54.08$), Block 2 ($F = 20.99$), Block 3 ($F = 17.80$), Block 4 ($F = 19.90$), Block 5 ($F = 15.83$), and Block 6 ($F = 4.71$). More random errors occurred in Cond. II and III than in I at the first block of trials ($F = 42.81$).

In order to facilitate comparisons between Cond. I, II, and III, mean percent errors for each type of error were computed separately for the strong and weak word pairs of each condition. This information is presented in Fig. 3.

DISCUSSION

The effect of increasing the number of competitive connections proved to be an interesting one. On the weak pairs, poorer performance was found to occur in Cond. I, which had two direct competitive connections, than in Cond. II and III, each having one such connection. Also, more response errors occurred in Cond. I, having two competitive connections, than in Cond. III, which had only an inter-word-pair S-R connection. The prediction that more generalized response errors would occur in learning the strong pairs in Cond. I and II than in III, though shown in Fig. 3 to be in the right direction, was not supported statistically. The failure, in general, to obtain differences among the strong S-R pairs may have been due to a partial masking of the effects of competition by the initially high associative strengths of these pairs. An inspection of Fig. 2 reveals that the initial performance level for the strong pairs was quite high.

A comparison of the two types of competitive connection, interstimulus item and inter-word-pair S-R, indicates that the inter-word-pair S-R connection proved to be a more effective source of interference than the interstimulus item connection. However, significantly more stimulus errors occurred on both the strong and weak word pairs in those conditions (I and II) having an interstimulus item connection than in the condition (III) not having such a connection. Also, more response errors occurred on the weak pairs in the conditions (I and III) having an inter-word-pair S-R connection than in the condition (II) not having such a connection.

Thus, the interstimulus item connection and the inter-word-pair S-R connection were both seen to be adequate sources of interference. They not only produced different numbers and kinds of errors but also seemed to have a greater effect in combination than individually. The presence of both types of interference in the same task would seem, therefore, to make a precise interpretation of the data rather difficult. Accordingly, an implication of these findings seems to be that the two new models (II and III) might prove less complicated psychologically than the original method for the purpose of assessing hypotheses concerning the interaction of habit and drive.

No attempt was made to assess the effects of varying the associative strengths of competitive connections, of including more than two competitive connections per set in the task, or of employing more than two S-R pairs per set. Logically, it would seem possible to construct from zero to four competitive connections between two given S-R pairs. These connections might be of the interstimulus item, inter-word-pair S-R, or interresponse item variety. Thus number and type of competitive connections as well as the strength of initial associations mediating both correct and incorrect response tendencies appear to be potentially important determinants of competition in this type of learning task.

REFERENCES

CASTANEDA, A., FAHEL, L. S., & ODOM, R. Associative characteristics of sixty-three adjectives and their relation to verbal paired-associate learning in children. *Child Develpm.*, 1961, **32**, 297–304.

RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

(Received February 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 169–175

TEMPORAL PROPERTIES OF REINFORCEMENT AFTEREFFECTS

E. J. CAPALDI AND LARRY R. STANLEY

University of Texas

Pattern running performance as a function of intertrial interval was investigated in 3 separate experiments which employed rats trained in a straight alley under alternating partial reinforcement. In the initial experiment speed of running on the nonreinforced trials late in training was equivalent for the 15-sec. and 20-min. intertrial interval groups, as it was for the 2-min. and 10-min. groups; however, the former groups ran more slowly on these trials than did the latter ones. Substantially similar results were obtained in the subsequent experiments. On the basis of these results the tenability of the Hull-Sheffield temporal assumptions was questioned and new temporal assumptions were advanced.

The assumptions of Hull (1952) and of Sheffield (1949) that aftereffects dissipate or decay as a function of time and that such stimuli do not become a portion of the total stimulus complex on an immediately subsequent trial if considerable trial spacing is employed appear to have found general acceptance among learning theorists (Lewis, 1960). However, there seem to be at least two reasons for subjecting these temporal assumptions to empirical test. The occurrence of increased resistance to extinction following partial reinforcement under widely spaced trials (e.g., Weinstock, 1954) does not necessarily infirm Hull's hypothesis; rather, these data could be interpreted to mean that aftereffects remain functional for a considerable period. Further, with the exception of the results from the Sheffield (1949) investigation, which have failed of confirmation on several occasions, the Hull-Sheffield temporal assumptions are not based on data available from the instrumental situation, where they are most often applied, but come to us from classical conditioning (Hull, 1952).

The attempt was made here to determine the temporal characteristics of aftereffects by employing alternating partial reinforcement under several conditions of trial spacing. A variety of massed trial studies (e.g., Capaldi, 1958; Tyler, Wortz, & Bitterman, 1953) have shown that under an alternating partial pattern rats eventually run relatively slowly on the nonreinforced trials, relatively rapidly on the reinforced ones. We assume that such pattern running reflects the control of the instrumental response by aftereffects. Hull's hypothesis holds that conditioning will not occur in the absence of reinforcement; accordingly, under an altermating pattern the aftereffect of nonreinforcement but not the aftereffect of reinforcement will acquire with training a considerable capacity to evoke the instrumental response. But these aftereffects are imbedded in a total stimulus complex which includes stimuli arising from the alley, the experimental room, etc. Thus, the instrumental response would tend to be strongly evoked by the stimulus complex which includes the aftereffect of nonreinforcement and less strongly evoked by the stimulus complex which includes the aftereffect of reinforcement. Presumably, it is because of the similarity between these stimulus complexes that pattern running requires a considerable number of trials for its appearance and that following quite extensive training the rat will nevertheless slowly traverse the runway on nonreinforced trials rather than stop running entirely (Bloom & Capaldi, 1961).

It should be clear, on the basis of the above analysis, that the development of pattern running under an alternating schedule at a given intertrial interval implies that the aftereffect is functional, and a failure of pattern running to develop, i.e., a tendency for running to be as rapid on nonreinforced as on reinforced trials, would suggest that the aftereffect is nonfunctional.

EXPERIMENT I

Method

Subjects.—The *S*s were 24 male and 24 female naive Sprague-Dawley strain rats about 90 days old at the beginning of the experiment from the colony maintained by the Department of Psychology, the University of Texas. Six males and 6 females were randomly assigned to each of four groups. Due to sickness or death, 5 *S*s were discarded during the experiment, 1 from the 15-sec. intertrial interval group and 2 each from the 2-min. and 10-min. groups. The data for these *S*s were discarded.

Apparatus.—The apparatus was a straight-alley runway which was a total of 72 in. long, 5 in. wide, enclosed by sides 8 in. high, constructed entirely of wood and painted a mid-gray throughout. An unpainted, sheet metal start-box door was located 12 in. from the proximal end of the alley and was so constructed that upon being raised 2 in. a Standard Electric timer, which measured in .01 sec., was automatically started. The clock was stopped when *S* interrupted a photocell beam located 48 in. from the start-box door. One inch beyond the photocell beam was located another sheet metal door which could be lowered so as to restrict *S* to the goal region. The end of the runway terminated in a 5 in. wide × 3 in. long "elbow" bend in which food could be placed. Two identical-appearing food compartments; i.e., elbow bends, were employed, one for reinforced trials, the other for nonreinforced trials. The elbow arrangement prevented *S* from determining whether the food compartment was baited until it had passed beyond the photocell beam and the goal-box door. The food compartments, mounted on runners, could be shifted into position noiselessly. The entire goal region was covered by hinged hardware cloth.

Procedure.—On each of the initial 10 days *S*s were deprived of food for 23 hr., handled in groups of 12 for approximately 1 hr., and returned to the home cage—each *S* being individually housed—where they were allowed to eat a wet mash for approximately 1 hr. On Days 11 and 12 each *S* was allowed individually to explore the runway for 5 min., no food being available and both doors being open.

Each *S* was given eight trials per day for 23 days, all trials being separated by the appropriate intertrial interval. Each *S* was taken from the home cage and deposited in the start box. Approximately 3 sec. later the start-box door was raised. When *S* entered the goal region *E* lowered the goal-box door. The *S* was confined to the goal region for 20 sec. following each run. The *S* was then taken to the home cage, which served as the place of confinement during the intertrial interval, to await further trials. On each day all even-numbered trials were reinforced with a wet mash, all odd-numbered trials, nonreinforced.

Four intertrial intervals were employed: 15 sec., 2 min., 10 min., and 20 min. Because of scheduling difficulties, the running period proper was divided into five segments. In the initial segment one-half the *S*s from the 15-sec. group were run, then in the second segment all *S*s of the 10-min. group were run. In the third segment all *S*s of the 2-min. group were run followed in the fourth segment by the remainder of the *S*s in the 15-sec. group. Finally, all *S*s in the 20-min. group were run. All trials were administered by a single *E* from 10 A.M. to 10 P.M. each day. A 2-hr. "break" occurred between the administration of trials to the 10-min. and the 2-min. groups. All *S*s were fed 1 hr. per day, *S*s in the 20-min. group being fed 10 min. following the last trial of the day. The *S*s in the 15-sec. group were fed for 1 hr. approximately 2½ hr. following the final trial of the day, and a similar procedure was followed for the remaining groups. Thus, at the beginning of the daily trials all *S*s had been deprived of food for approximately 20½ hr. Throughout all phases of the experiment *S* was allowed a maximum of 60 sec. to traverse the runway. If *S* did not traverse the runway within the arbitrary time limit, it was encouraged to do so by gentle pushing, etc. The intertrial intervals given above are lower values. For example, a 20-min. *S* was never run following an interval of less than 20 min. and on occasion the interval might be as great as 21 min. For a 15-sec. *S*, as much as 18 sec. might elapse between some trials, and so on.

Results

In Fig. 1 is plotted for each of the groups on each day of acquisition the mean of the log times on reinforced and on nonreinforced trials. On the initial days and for some days thereafter, with exceptions that can be noted by examining the figure, each group ran more rapidly on its nonreinforced trials; i.e., following reinforced trials, than on its reinforced trials. The differences were significant on Day 1 for the 2-min., $F(1, 9) = 10.87$, $p < .01$; the 10-min., $F(1, 9) = 5.58$, $p < .05$; and the 20-min., $F(1, 11) = 7.70$, $p < .05$ groups; however, not until Day 3 was the difference significant in the case of the 15-sec. group, $F(1, 10) = 10.58$, $p < .01$. Late in training each group ran more slowly on its nonreinforced trials. The initial day on which running was significantly slower on the nonreinforced as opposed to the reinforced trials for the 15-sec. group was Day 18 ($F = 11.16$, $p < .01$), for the 2-min. group Day 18 ($F = 10.85$, $p < .01$), for the 10-min. group Day 17 ($F = 11.20$, $p < .01$), and for the 20-min. group Day 17 ($F = 6.59$, $p < .05$). For each group all subsequent daily differences between reinforced and nonreinforced trials were also significant. On the basis of these results it is apparent that aftereffects remain functional for at least 20 min. In addition, all groups exhibited pattern running following about the same number of training trials. Thus, it would appear on the basis of the Trials measure that aftereffects do not undergo diminution within a 20-min. period.

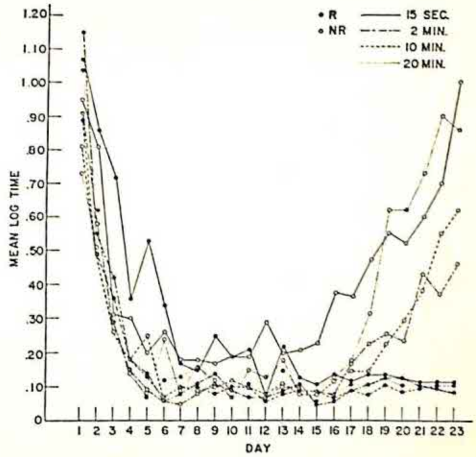


FIG. 1. Mean of log running times on reinforced and on nonreinforced trials on each of the acquisition days for each of the groups.

Statistical tests involving comparisons between the several groups were also made and are reported below. The effect of intertrial interval on performance in the early trials is, of course, of less interest to this investigation than is performance on the later trials. Nevertheless, early trial data were analyzed statistically. An analysis of variance on the mean of the log times on Days 1 through 5 indicated that significant differences occurred between the groups on reinforced, $F(3, 39) = 3.05$, $p < .05$, but not on nonreinforced trials ($F = 1.40$). It would appear that in the early trials the effect of longer intertrial intervals in producing faster running is greater following nonreinforced than following reinforced trials. Differences between the groups on nonreinforced trials were significant on the Day 6 through 13 block of trials ($F = 3.97$, $p < .05$) and on the Day 14 through 18 block of trials ($F = 3.83$, $p < .05$). These differences are not meaningful from the viewpoint of an aftereffects analysis, however, since in none of the groups had pattern running developed to any appreciable extent. In addition, differences on the reinforced trials were significant (for Days 6 through 13, $F = 6.02$, $p < .01$; Days 14 through 18, $F = 3.53$, $p < .05$); thus, the nonreinforced trial differences through Day 18 probably reflect slower running per se to as great or perhaps to a greater extent than they reflect the influence of the aftereffect.

Differences on the Day 19 through 23 block of trials were not significant on reinforced trials ($F < 1$) while differences on the nonreinforced trials were significant ($F = 3.44$, $p < .05$). Since pattern running had developed in this block of trials and differences on reinforced trials were not significant, the nonreinforced trial differences can be taken to reflect the influence of the aftereffect. An extension of Duncan's multiple range test to group means with unequal numbers of replications (Kramer, 1956) was employed using these nonreinforced trial data. The difference between the 15-sec. and the 20-min. groups was not significant nor was the difference between the 2-min. and the 10-min. groups. The 15-sec. group, as did the 20-min. group, differed at slightly beyond the .05 level from the 2-min. and from the 10-min. group.

EXPERIMENTS II AND III

The differences on nonreinforced trials late in training could have been due to (*a*) sampling; (*b*) an extraneous or uncontrolled variable which resulted in atypically slow running on the nonreinforced trials late in training under the 20-min. condition, e.g., the 20-min. group was run each day during the final hours of the experiment; (*c*) an uncontrolled variable which contributed to faster running on nonreinforced trials under the 2-min. and 10-min. conditions; (*d*) some variable exclusively associated with the 2-min. and 10-min. intervals but not the 15-sec. and 20-min intervals, e.g., greater excitement or agitation. This latter alternative suggests that a repetition of Exp. I, even under the most ideal experimental conditions, would produce substantially similar results. Actually, casual observation indicated that the 2-min. and 10-min. *S*s were more agitated (greater incidence of biting, struggling, defecation, etc.) than the 15-sec. and 20-min. *S*s.

Experiment I was replicated in part; in Exp. II a 15-sec. and a 2-min. group were trained and in Exp. III a 15-sec. and a 20-min. group were trained. The experimental procedures employed were identical to those previously employed except for the changes noted below. Some changes represent an improvement in experimental technique made possible by the less demanding physical requirements of the present experiments, e.g., the daily running order of the groups was better balanced. Other changes were introduced in an attempt to facilitate the appearance of slow running on the nonreinforced trials, e.g., opening of the start door seemed to function as a potent stimulus for leaving the start area; thus the door was removed.

Method

Subjects.—The *S*s were, in Exp. II, 21 female and, in Exp. III, 20 male colony rats about 90 days old at the start of the experiment.

Apparatus.—The same apparatus was employed except for the following modifications. The start door was removed. The clock was now started by placing *S* on a 5 × 8 in. platform or floor treadle suspended over a microswitch. The beginning of the floor treadle was located 2 in. from the extreme proximal portion of the alley. The photocell was moved 8 in. further from the alley's origin. An additional change occurred in Exp. III; the alley was covered throughout its length by a hinged, plastic top.

Procedure.—There were 12 trials per day for 18 days in Exp. II and 10 trials per day for 14 days in Exp. III. As before, the initial daily trial was reinforced, etc. The initial 15 days comprise the acquisition phase of Exp. II, the remaining 3 days the shift phase. The group which had been trained under the 15-sec. intertrial interval was trained on Days 16, 17, and 18 under the 2-min. interval, while the 2-min. group was shifted to the 15-sec. interval on these days. In Exp. III, the 15-sec. group was shifted to the 20-min. intertrial interval and the 20-min. group to the 15-sec. intertrial interval on Days 13 and 14. There were 10 *Ss* in each of the groups except the 2-min. group of Exp. II which contained 11 *Ss*. However, an *S* in the 2-min. group died on Day 12 of training. The data for that *S* were discarded. The daily running order in Exp. II was as follows: a 15-sec. *S* was run followed by 3 2-min. *Ss*, then 5 15-sec. *Ss*, then 6 2-min. *Ss*, 4 15-sec. *Ss*, and, finally, 2 2-min. *Ss*. In Exp. III, *Ss* were run as follows: each *S* in the 20-min. group received the initial trial in turn and then a 15-sec. *S* was given its 10 runway trials. Then the second trial was administered to each of the 20-min. *Ss*, followed again by a 15-sec. *S* receiving all its daily trials, and so on.

Results

The results of Exp. II are presented in Fig. 2. Differences between the groups on reinforced and on nonreinforced trials were not significant over the initial 5 days of training. On Days 6 through 10 reinforced trial differences approached significance, $F(1, 18) = 3.09$, $p < .10$, but nonreinforced trial differences did not ($F = 1.95$). Again, it appears that in the earlier stages of training intertrial interval exerts a greater effect foll ing nonreinforced than following re forced trials.

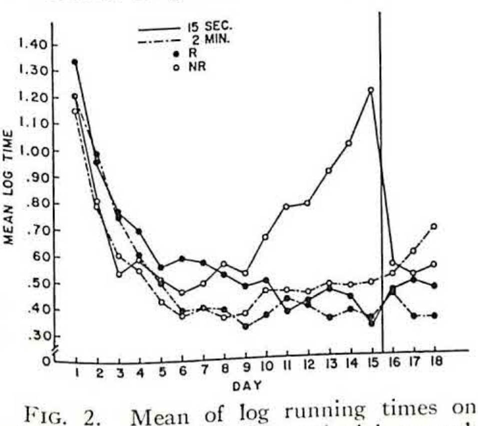


FIG. 2. Mean of log running times on reinforced and on nonreinforced trials on each of the acquisition days for each group.

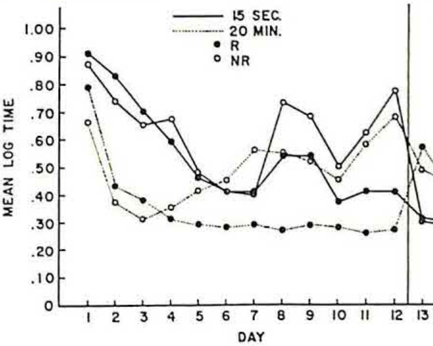


FIG. 3. Mean of log running times reinforced and on nonreinforced trials each of the acquisition days for each grou

The initial day on which sigr cantly slower running on the n reinforced as opposed to the reinforc trials occurred was, for the 15-s group, Day 11, $F(1, 9) = 9.71, p <$. and, for the 2-min. group, Day $F(1, 9) = 10.47, p < .05$. Differen on each subsequent acquisition c were significant in the case of 15-sec. group. However, for 2-min. group differences on Days 12, and 14 were not significant, dicating unstable pattern running.

Differences between the grou on the reinforced trials of Da 11 through 15 were not signific ($F < 1$). However, differences on nonreinforced trials were highly s nificant ($F = 23.67$, $p < .001$).

On Day 16, the shift day, runn continued to be slower on nonre forced as opposed to reinforced tr for each group. For the 15-sec. gro the difference was not significant any of the shift days. However, 2-min. group appeared to fare qu well on the shorter interval, difference on Day 16 being signific beyond the .10 level, $F(1, 9) =$ 3 and the difference on Day 17 and Day 18 being significant beyond .05 level ($F = 9.13$ and 9.28, spectively). It appears from the figure that both groups performed about equally well on the initial shift day, but, as the statistical analyses reported above indicate, they did not. On the initial shift day only 1 *S* in the 2-min. group ran more slowly on its reinforced as opposed to its nonreinforced trials, while in the 15-sec. group 2 *S*s ran more slowly on the reinforced trials, a third *S* running equally rapidly on both types of trials.

The results of Exp. III are presented in Fig. 3. Differences between the groups on reinforced, $F(1, 18) = 20.01$, $p < .001$ and on nonreinforced ($F = 11.61$, $p < .001$) trials over the initial 4 days of training were significant. On Days 5 through 9 differences between the reinforced trials were significant ($F = 19.22$, $p < .001$) while differences on the nonreinforced trials were not ($F < 1$). It can be seen that pattern running developed relatively quickly, especially in the 20-min. group. The earliest day on which significantly slower running occurred on the nonreinforced as opposed to the reinforced trials was, for the 15-sec. group, Day 10, $F(1, 9) = 5.66$, $p < .05$; and, for the 20-min. group, Day 7, $F(1, 9) = 6.32$, $p < .05$. For both groups, all subsequent daily differences were significant. Analysis of the running times of Days 10 through 12 indicated that differences between the groups on reinforced trials were significant, $F(1, 18) = 17.89$, $p < .001$ while differences on the nonreinforced trials were not significant ($F < 1$). Since on these days the 20-min. group ran more rapidly than the 15-sec. group on the reinforced trials and equally slowly on the nonreinforced ones, the notion of a nonfunctional or diminished aftereffect hardly received support. In general, the results of Exp. II and III are in agreement with the results given by the corresponding groups in Exp. I.

It will be noted in Fig. 3 that for both groups pattern running was completely disrupted by the change in intertrial interval. Apparently, pattern running will not—at least initially—survive a relatively large shift in intertrial interval regardless of the direction of shift, i.e., from longer to shorter, or vice versa. Interestingly, the former 15-sec. group ran faster than the former 20-min. group on each of the shift days. Analysis of the log means of the shift days indicated that the differences between the groups were significant on reinforced—$F(1, 18) = 14.86, p < .01$—and on nonreinforced ($F = 8.80$, $p < .01$) trials.

DISCUSSION

It appears from these results that aftereffects remain functional for at least 20 min. The conclusion is unambiguously supported by the occurrence of pattern running in a 20-min. group in Exp. I and III. In addition, it appears that aftereffects fail to undergo diminution within 20 min., notwithstanding the performance of the 2-min. and 10-min. groups. The clearest and most obvious source of support for this conclusion is that in two separate experiments a 20-min. group ran as slowly on the nonreinforced trials late in training as a 15-sec. group. Accordingly, it is suggested that some variable (e.g., emotional excitement) contributed to relatively rapid running on nonreinforced trials at the 2-min. and 10-min. intervals. The shift data of Exp. II support this conclusion. The disruption shown by the shifted 15-sec. group cannot be exclusively explained in terms of generalization decrement occasioned by the introduction of a stimulus trace of a different temporal age because the shifted 2-min. group experienced a stimulus change of similar magnitude in the opposite temporal direction yet exhibited stable pattern running. Nor does an explanation of the performance of the shifted 2-min. group in terms of the opportunity for learning or conditioning to take place at the 15-sec. interval seem adequate. This, because recovery from disruption and the appearance of stable pattern running occurred following relatively few training trials at the new interval. However, the shift data in their totality are consistent with the view that both the 15-sec. and 2-min. groups had learned equally well but there operates at the 2-min. interval (and by implication the 10-min. interval) some variable which produces faster running on the nonreinforced trials.

The present results appear to support a conception of the temporal properties of aftereffects unlike that offered by Hull or Sheffield. First, it seems reasonable to abandon the dissipation assumption, assuming instead that aftereffects persist or, to use more neutral terminology, remain functional until such time as a new internal stimulus is substituted for the original one. It seems apparent that pattern running in the 20-min. group cannot be easily reconciled with the view that dissipation occurs as a mere function of time. On the other hand, available evidence indicates that aftereffects are subject to interference (Capaldi, Hart, & Stanley, 1963). Capaldi et al. have presented evidence which indicates that placing *S* in the baited goal box for a relatively brief period during the intertrial interval substitutes the aftereffect of reinforcement for that of nonreinforcement. The intertrial interval was 30 sec. Thus, their study indicates that under appropriate circumstances an aftereffect which would otherwise have remained functional for at least 20 min. can be interfered with in as brief a period as 30 sec. Capaldi et al. hypothesized that the extent of interference is probably related to the stimulus simularity which obtains between the original reinforcement situation which occasions the aftereffect on the one hand and the interfering situation on the other. If this assumption is valid, then results such as those supplied by Weinstock (1954) can be interpreted in terms of aftereffects. That is, it might be otherwise thought that when trials are separated by 24 hr., the receipt of the maintenance diet, occurring as it does following the daily runway trial, would serve to interfere with the aftereffects of nonreinforcement. However, it is apparent that since the maintenance diet is administered in the presence of stimuli, i.e., the home cage, quite unlike those provided by the goal box in which reinforcement and nonreinforcement occur, very little actual basis exists for interference. Consequently, an aftereffects hypothesis modified by the temporal assumptions presented here would expect increased resistance following partial training when trials are spaced at long intervals and the maintenance diet is administered in the presence of stimuli unlike those present on nonreinforced trials.

REFERENCES

BLOOM, J. M., & CAPALDI, E. J. The behavior of rats in relation to complex patterns of partial reinforcement. *J. comp. physiol. Psychol.*, 1961, **54,** 261–265.

CAPALDI, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, **51,** 367–371.

CAPALDI, E. J., HART, D., & STANLEY, L. R. Effect of intertrial reinforcement on the aftereffect of nonreinforcement and resistance to extinction. *J. exp. Psychol.*, 1963, **65,** 70–74.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 1956, **12,** 307–310.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57,** 1–28.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39,** 511–526.

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *Amer. J. Psychol.*, 1953, **66,** 57–65.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47,** 318–322.

(Received February 19, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 176–181

GENERALIZATION: I. GENERALIZATION GRADIENTS FROM SINGLE AND MULTIPLE STIMULUS POINTS. II. GENERALIZATION OF INHIBITION[1]

HARRY I. KALISH

State University of New York, Long Island Center

AND AUDREY HABER

Adelphi College

35 pigeons were trained to peck at a monochromatic stimulus of 550 mμ. Ss were then assigned to a group and extinguished to one of the following stimuli: 550, 540, 530, 520, 510, and 490 mμ. on each of 2 successive days. Gradients of generalization (single stimulus) from the mean total number of responses during extinction, were contrasted with a comparable gradient from Guttman and Kalish (1956) in which Ss sampled over the entire range of stimuli (multiple stimulus). The gradients showed marked differences which were discussed. Postextinction gradients were obtained for the 2nd part of the study by permitting Ss to sample over the range of stimuli from 510 to 550 in 10-mμ. steps (including 490 mμ.). The obtained gradients were compared with gradients generated from a "mirror-image" hypothesis of generalized inhibition. Certain marked discrepancies were observed between the predicted and obtained gradients which were supported by other studies of generalization of inhibition.

The present study consists of two parts: the first provides additional information concerning the relationship between generalization and discrimination from a comparison of primary stimulus generalization gradients derived from testing to single and multiple points on the spectral wave-length continuum. The second explores the postextinction gradient and the problem of generalization of inhibition. A number of studies (Grant & Schiller, 1953; Grice & Saltz, 1950; Kling, 1952; Margolius, 1955) have demonstrated that gradients of generalization can be recovered by relating number of responses during extinction to stimulus position on the continuum, and that, therefore, the presence of a gradient is not altogether contingent on *S*s sampling along the continuum.

While it is important to demonstrate that gradients of generalization can be obtained from the single stimulus situation in order to substantiate the argument that the generalization decrement is fundamental, rather than epiphenomenal, it is equally important to examine the differences in gradients which derive from different generalization testing procedures to determine whether these differences are, in turn, related to other factors such as discrimination. A comparison of generalization testing procedures might yield information concerning the relative proportion of responding attributable to discrimination and generalization.

Part I of the present study examines these differences by comparing gradients obtained from comparable *S*s

---

[1] This study was supported by Grant M-1459 from the National Institute of Mental Health, United States Public Health Service.

under single and multiple testing conditions.

Part II, a derivative of the method involved in Part I, makes it possible to observe the effects of extinction to a single stimulus on the subsequent postextinction gradient and to draw some inferences regarding generalization of inhibition.

METHOD

Part I

Apparatus.—An automatic key pecking apparatus similar to that described by Guttman and Kalish (1956) was used. The light source for illumination of *S*'s key was a Cambridge Thermionic Corporation monochromator which has been fully described elsewhere (Guttman & Kalish, 1956). No attempt was made to equate for apparent brightness along the spectral wave-length continuum since other studies (Guttman & Kalish, 1956; Honig, Thomas, & Guttman, 1959) have demonstrated that luminosity differences exert a negligible influence on the formation of spectral generalization gradients. Moreover, a gradient was generated for only one specific wave length and no attempt was made to compare gradients obtained from other regions on the spectrum.

Subjects and procedure.—Training—35 experimentally naive, white, Carneau pigeons were reduced to 80% of their ad libitum weight and trained by the method of successive approximations to peck at a plastic disc illuminated by monochromatic light of 550 mμ. When *S*s responded to the key and food magazine they were given 50 continuous 3-sec. reinforced trials and placed on a VI schedule (1-min. mean interreinforcement interval) during which 60-sec. stimulus-on intervals were alternated with 10-sec. stimulus-off intervals in order to facilitate changing of stimuli during generalization testing. During stimulus-on periods the key was illuminated by 550 mμ. and the box by a 25-w. bulb set beneath a sheet of translucent plastic which comprised the floor of the apparatus. In the stimulus-off periods both the bulb and stimulus source were off. When a criterion of approximately 1500 responses per 30-min. period was reached, *S*s were assigned to one of five extinction groups.

The criterion employed for assigning *S*s to an extinction group was based on the number of days of training prior to the attainment of 1500 responses per 30-min. period. This method of assignment was ultimately used because the simple criterion of response rate did not adequately equate *S*s when the extinction data were examined. Instead, number of responses during extinction appeared to be a complex curvilinear function of number of days of training prior to reaching the criterion. An attempt was made, therefore, to equate *S*s on number of days of training. With the exception of 530 all of the groups appear to be within reasonable range and there is no relationship between the means and *SD*s or between either measure and the order of the extinction group.

Extinction.—Extinction followed the day after the training criterion was reached and each *S* in a given extinction group was administered 132, 30-sec. nonreinforced trials to one of the following stimuli: 550, 540, 530, 520, 510, and 490 mμ. on each of 2 successive days. Two days of extinction were introduced to minimize the effects of spontaneous recovery during the generalization testing. Thirty-second stimulus-on periods were alternated with 10-sec. blackouts and number of responses recorded during the stimulus-on periods.

Part II

Generalization testing.—Generalization tests were administered the day following the second extinction series. Six different stimuli (550, 540, 530, 520, 510, and 490 mμ.) were randomized within a series and 12 different random series were presented to each *S* resulting in a schedule of 72 stimulus presentations. Each nonreinforced stimulus presentation was 30 sec. followed by a 10-sec. stimulus-off interval during which the wave-length value was changed. An additional group of 4 *S*s extinguished on 540 was tested on 560, 550, 540, 530, 520, and 510 mμ. in order to compare the decrement at 540 with that at 560 mμ.

RESULTS

Single and Multiple Generalization Gradients

The analysis most directly relevant to the first part of the present study is the comparison of generalization gradients shown in Fig. 1. The gradient represented by a solid line was obtained from a previous study by Guttman and Kalish (1956) in

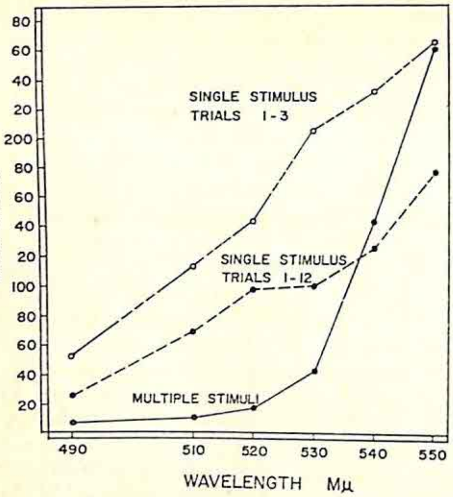


FIG. 1. Single and multiple stimulus gradients.

which *S*s were trained to 550 mμ. and tested over the range of values from 490 to 610 mμ. in 10-mμ. steps with the exception of the 490–510 and 590–610 differences. This gradient has been labeled "Multiple Stimuli" because the same *S* is represented at each stimulus value, having sampled the entire range of stimuli during generalization testing. By contrast, the gradient labeled "Single Stimulus Trials 1–12" was derived from the present study in which *S*s were exposed to only one stimulus on the continuum during extinction. The values for this gradient represent mean total number of responses for Extinction Trials 1–12.

Since the values for the multiple stimulus gradient (Guttman & Kalish, 1956) represent mean number of responses per 6-min. period, the mean of the 12 extinction trials for the extinction groups shown in Fig. 1 was further divided by 11 to yield a comparable response rate. These values are plotted as the single stimulus gradient in Fig. 1. When the two gradients are compared certain striking differences become apparent. The multiple stimuli gradient is markedly concave and the difference in responses between 550 and the remaining stimuli are quite large. The single stimulus gradient, on the other hand, is predominantly linear since the only significant F in an analysis of orthogonal polynomials was for the linear component, $F(1, 29) = 20.42, p < .01$. The same analysis for the multiple stimulus gradient revealed both a significant linear and quadratic trend, $F(1, 25) = 111.33$, $p < .01$, and $F(1, 25) = 28.80$, $p < .01$, respectively.

These differences between both gradients were further substantiated by an analysis of goodness of fit using the multiple gradient as the hypothetical gradient. That $F(5, 34) = 4.73$, ($p < .01$) indicates that the two gradients are markedly different in form.

The third gradient in Fig. 1, labeled Single Stimulus Trials 1–3, represents

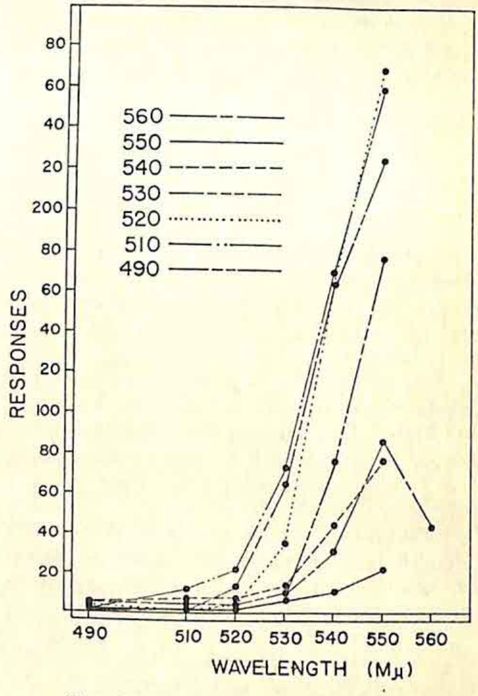


FIG. 2. Postextinction gradients.

the mean of the first three extinction trials for the individual extinction groups. The values in this gradient are part of the total response single stimulus gradient shown in Fig. 1. The data were fractionated in this manner to determine whether the form of the total single stimulus gradient is an artifact of averaging or a phenomenon which is preserved throughout the entire extinction series. The fractionated single stimulus gradient also appears to be linear when tested by the method of orthogonal polynomials, $F(1, 29) = 22.81$, ($p < .01$); for all other polynomials $F < 1.00$.

Generalization of Inhibition

The gradients in Fig. 2 are the result of generalization testing after the second day of extinction. These are multiple stimulus gradients, since the groups assigned to the particular wave-length values during extinction were subsequently tested over the range of values from 490 to 550 mμ. (with the exception of the group labeled 560 which was tested over the range 510–560 mμ.).

The effect of the extinction process on the subsequent formation of a generalization gradient appears to have been negligible for the groups extinguished at 490, 510, and 520 mμ. Moreover, these gradients bear a striking resemblance to the 550 gradient (Fig. 1) from the Guttman and Kalish (1956) study which indicates the reliability of the procedures used to generate the gradients in these studies of generalization. The remaining gradients (530, 540, 550, and 560 mμ.), on the other hand, demonstrate a decrement in response rate which appears to be influenced by the extinction procedure and reflects a process analagous to generalized inhibition.

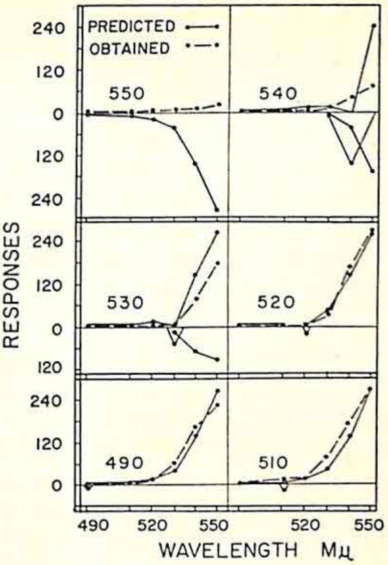


FIG. 3. Comparison of predicted and obtained gradients of inhibition and excitation.

An attempt was made to account for the decrements observed in Fig. 2 by generating the hypothetical gradients in Fig. 3. These gradients were derived by using the 550-mμ. excitation gradient from the Guttman and Kalish (1956) study as a model for the inhibition gradient generated at each point of extinction. The height of the inhibition gradient was assumed to be equivalent to the height of the 550-excitation gradient at that point. These are the bidirectional gradients contained in the lower halves of Fig. 3 and, it is interesting to observe that, following these assumptions, the only gradients which appear to be affected by the gradients of inhibition are 550, 540, and 530 mμ. The predictions for the remaining gradients 520, 510, and 490 mμ. appear to be trivial, since these would affect the gradient only at the point of extinction and with a reduction in response which is probably not greater than the amount of error involved at that point.

The discrepancies between the hypothetical and obtained gradients are contained in Fig. 3. These comparisons indicate no difference between the hypothetical and obtained gradients for 490, 510, and 520 as predicted. In each case, the F for goodness of fit of the obtained to the predicted gradient is less than 1. By contrast the test of goodness of fit yielded $F(5, 25) = 3.01$, $p < .05$ for 530 and $F(5, 25) = 36.99$, $p < .01$ for the 540 comparison. A residual gradient appears to have been generated for the *S*s trained and extinguished at 500 since the predicted gradient for this group should have been zero. Also, the decrement in the residual gradient appears to be significant since the $F(5, 20) = 3.45$ is equivalent to a $p = .025$.

The unidirectional curves for 530 and 540 mμ. in the lower halves of Fig. 3 are gradients of inhibition obtained by calculating the difference between the expected and obtained gradients of excitation using the 550 group from the Guttman and Kalish (1956) study for comparison purposes. It is apparent from Fig. 3 that the maximum decrement for the gradient at 530 and 540 mμ., calculated in this manner, occurs at the training stimulus instead of at the point of extinction.

DISCUSSION

Multiple and Single Stimulus Gradients

The results from the present study suggest that the gradient which derives from sampling over the range of a specific continuum differs in certain salient respects from that which is generated when S is exposed to only one stimulus. The concave appearance of the multiple stimuli gradient in Fig. 2, for example, appears to be the result of an increase in response to the training stimulus and a decrease in the adjacent stimuli when compared to the single stimulus gradient. In effect, the heightened response at 550 mμ. for the multiple stimulus gradient may reflect a differentiating process which increases the response to the training stimulus in relation to the other stimuli throughout the course of generalization testing. Similar results have also been observed in other studies of generalization. When number of responses for each stimulus value is calculated as a percentage of the total (relative generalization) for any particular test series in the Guttman and Kalish (1956) study, percentage of responses to the training stimulus increases with successive generalization test trials and decreases for the adjacent stimuli. Other studies have indicated the extent to which the generalization process can be altered by procedure in testing and training. Reinhold and Perkins (1955) have obtained evidence indicating that the gradient for one dimension can be altered by training on a different dimension. Specifically, the effect of training was to steepen the gradient for the second dimension, a result which the *E*s attributed to the development of a set to discriminate analagous to the development of learning sets.

Apparently, some considerable part of the decrement in the gradient appears to be an artifact of the generalization testing procedures involving a differentiating process.

Generalization of Inhibition

The results obtained in the second part of the present study indicate that, while certain evidence concerning generalization of inhibition is corroborated, other views may need revision. The results obtained for the group trained and extinguished at 550 mμ., for example, are substantially similar to those obtained by Hovland (1937), Honig (1961), and Kling (1952). All of these studies indicate that when S is trained to all stimuli (in the case of Kling to the two extreme stimuli) and extinguished to one of the extremes, the derived gradients of inhibition resemble the gradients of excitation in that the reduction in response is inversely related to distance from the point of extinction.

The residual gradient for the 550-mμ. group, (Fig. 2 and 4) suggests that extinction was not complete. When the difference between the residual gradient and the comparison 550-mμ. group is calculated, the resulting gradient of inhibition is similar in form to the gradient of excitation; a finding which confirms previous studies.

The same situation does not prevail, however, when a stimulus other than the training stimulus is extinguished. The gradients of inhibition in this case do not appear to be mirror images of the excitation gradient and the inhibitory effects are restricted to a region near the training stimulus. In the present study this region extended to 530 mμ.; beyond this, extinction at any point did not seem to affect the gradient in any meaningful manner. The gradients which are formed by the groups extinguished at 490, 510, and 520 mμ. were almost identical with the 550 comparison gradient from the Guttman and Kalish (1956) study. The postextinction gradients for 530, 540, and 550 mμ. in Fig. 2 appear to conform to the expected gradients for these groups since they are arranged in an approximate sequential order beginning with 550 as the lowest and 530 mμ. as the highest. When the 550 gradient from Guttman and Kalish (1956) is used as a basis for computing the decrement, however, the gradients of inhibition necessary to account for the postextinction gradients are not symmetrical around the point of extinction. Instead, both the 530 and 540 inhibition gradients exhibit a common characteristic which places the point of greatest decrement at the training stimulus. The group of *S*s trained at 540, extinguished at 540, and tested over the range from 510 and 560 mμ. appears to be similar to the group extinguished at 540 and tested over the range from 490 to 550 mμ. For this group the greatest relative decrement also appears at the training stimulus and the decrement at 560 appears to be similar to that for 540 mμ.

While it is difficult at the present time to account for these findings they are, nevertheless, supported by Hanson's (1959) and Honig, Thomas, and Guttman's (1959) studies. In the Honig, Thomas, and Guttman (1959) experiment gradients of inhibition derived from differences between the postextinction gradient and a control gradient for *S*s trained at 550 and extinguished at 570 do not have their peak at 570 mμ. In Hanson's (1959) study the point of greatest decrement also shifts from the point of extinction toward the training stimulus.

REFERENCES

GRANT, D. A., & SCHILLER, J. J. Generalization of the conditioned galvanic skin response to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.

GRICE, S. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HANSON, H. H. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58**, 321–334.

HONIG, W. K. Generalization of extinction on the spectral continuum. *Psychol. Rec.*, 1961, **11**, 269–278.

HONIG, W. K., THOMAS, D. R., & GUTTMAN, N. Differential effects of continuous extinction and discrimination training on the generalization gradient. *J. exp. Psychol.*, 1959, **58**, 145–152.

HOVLAND, C. I. The generalization of conditioned responses: The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

KLING, J. W. Generalization of extinction of an instrumental response to stimuli varying on the size dimension. *J. exp. Psychol.*, 1952, **44**, 339–346.

MARGOLIUS, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.

REINHOLD, O. B., & PERKINS, C. C., JR. Stimulus generalization following different methods of training. *J. exp. Psychol.*, 1955, **49**, 423–427.

(Received March 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 182–189

EFFECTS OF INCENTIVE MAGNITUDE ON RUNNING SPEEDS WITHOUT COMPETING RESPONSES IN ACQUISITION AND EXTINCTION[1]

MELVIN H. MARX AND AARON J. BROWNSTEIN[2]

University of Missouri

32 albino rats were given 1 trial daily in a 14-ft. runway for a total of 73 acquisition and 72 extinction trials. Trials with competing responses (stopping and/or reversing and retracing) were scored separately. Except for the slow running of the 8% sucrose group, running speed with gross competing responses thus eliminated did not consistently vary with percentage of sucrose reinforcement. However, it increased regularly during acquisition and decreased regularly during extinction. The data also indicate the role of competing responses to be markedly greater, in slowing performance, during extinction. It is concluded that changes in running can be directly produced by affecting the vigor of responding, and are not merely effected indirectly, by changes in the amount of interfering behavior.

Recent experimental work has provided a new type of data bearing on the question of the extent to which certain motivational variables operate directly, through variations in the vigor of behavior, or indirectly, through variations in the amount of competing behavior. Thus Cotton (1953) controlled competing responses in the runway by measuring trials on which such occurred separately from those on which none occurred. All of the competing responses observed by Cotton required the rat *S* to stop, and so were incompatible with forward running (e.g., face-washing, reversing in the runway, etc.). On trials on which no stopping occurred, he found statistically reliable but greatly reduced variations in running speed as a function of hours of food deprivation. Pereboom and Crawford (1958) manipulated amount of reward (10-mg. vs. 50-mg. pellets) and found statistically reliable but relatively slight changes over trials, when only forward progress in the runway was measured. A later runway experiment by King (1959) varying hours of food deprivation indicated much more marked variations in speed without competing responses than Cotton had earlier found.

Further experimental data on this problem are needed, especially on rate of approach to asymptote and extinction. Although both Cotton and Pereboom and Crawford reported statistically reliable differences, with competing responses controlled, there is some question in each case as to whether these differences were artifactual. Cotton's low-deprivation groups produced an appreciably greater number of trials on which times could not be obtained because of the failure of the timing device. He conjectured that these trials were mostly very fast ones on which the animals jumped over the photoelectric beam. If this conjecture is accurate the direct-running times for Cotton's low-deprivation conditions were spuriously slow, and the statistical reliability of his drive-behavior relationship is questionable. Pereboom and

---

[1] This research was supported by United States Public Health Service Grant M-817 from the National Institute of Mental Health, Bethesda, Maryland.

[2] Now at Florida State University.

Crawford used an admittedly faulty direct-run measure. They included all of the forward running done by the rat, retracing as well as original progress. Thus their early trials, if these involved more retracing, were also spuriously slow, and the statistical reliability of their trials effect is suspect.

The present research was designed to produce data on the course of running speed over acquisition and extinction with competing responses eliminated.[3] Incentive magnitude was varied in the experiment in order to examine its relationship to running speed with competing responses eliminated. Four concentrations of sucrose in water were used as incentives. A single daily trial was given in both acquisition and extinction, and two measures of running performance without competing responses were used: forward-progress (FP) speed (only time consumed in actual forward progress through runway counted; all stopping, reversing, and retracing of runway in forward direction excluded) and direct-run (DR) speed (only those trials in which no stopping or reversing and retracing was observed). The FP measure has the advantage of being available on every trial; the DR measure has the advantage of being more unambiguously free of competing responses, and more reliable (less problem of *O* judgment and no complication from possible variations in progress due to differential speeds associated with stopping and starting).

METHOD

Apparatus

The runway used was 14 ft. long, 3¼ in. wide, and 5½ in. high. It had a 1-ft. start box and a 16-in. end box, each the same width and height as the stem. The masonite sides of the runway were encased in a soundproof housing. Aluminum grids served as floor switches for timing through Standard Electric timers. A Hunter decade timer provided automatic opening of the start door, 3 sec. after the rat was placed in the start box and the top closed. The first floor switch was set 6 in. past the start door, to measure starting latency. The final floor switch was set 8 in. behind the end-box door, and timed running in the 14-ft. stem. A burette containing the sucrose-solution incentive was placed in the top of the end box, at the far end, 2 in. from the floor.

Although this runway is normally used with a sound-insulated cover, for present purposes the cover was replaced with a hardware cloth top in order that an *O* could watch the forward progress of the rat. Thirteen 7½-w. light bulbs were placed at 1-ft. intervals on the outside of the runway, where they could not influence *S*, so that *E* could mark the point of retrace by turning on the nearest bulb. The only light visible in the runway came from two overhead fixtures, 6 ft. above the runway and 4 ft. to the side. Each had one 100-w. shielded bulb, and there was uniform illumination in the runway. A ventilating fan ran continuously and served as a noise mask.

Subjects

Thirty-two female Sprague-Dawley albino rats, obtained from the Holtzman Company, were the *S*s. They had been previously used in a bar press experiment involving licking of water and 32% sucrose, but had no runway experience. They were approximately 120 days of age at the start of the experiment. They were randomly assigned to four incentive groups, receiving 8, 16, 32, or 64% sucrose, by weight in water, during acquisition.

---

[3] After the present experiment was completed and this report prepared, two more recent experiments appeared with relevant data. Cicala (1961) reported two experiments in which he found the usual relationship between running speed and hours of deprivation to persist with little change when only direct-running trials were measured. His methodology was somewhat similar to that of King (1959) in that a short (4-ft.) runway was used and several trials per day were run.

The second of the recent reports is by Kintsch (1962). He gave 55 acquisition trials, 1 per day, in a 7-ft. runway. Hours of water deprivation and amount of water reinforcement were both positively and reliably related to starting and running speeds. Very few competing-response trials occurred, especially in the later stages of acquisition. Kintsch concludes that the differences between groups, and also the improvement in speeds over trials, were not artifacts of competing responses.

Procedure

Accommodation.—All *S*s were placed on a 23-hr. food deprivation schedule for 7 days before experimentation started. They were fed Gaines dog pellets for 1 hr. daily. Water was available ad libitum.

Magazine training.—Four days of magazine training were given. Each *S* was placed individually in the end box, with the door closed, and permitted 1 min. licking of the appropriate sucrose solution from the burette. Timing was begun as soon as licking started and continued regardless of pauses.

Acquisition.—One trial per day for 73 days was given. Licking of the sucrose solution for 20 sec. was permitted, timed from the initial lick. The *S* was then removed from the end box and returned to his home cage.

Four time measures were obtained:

1. *Starting time* was measured by a Standard Electric timer that began when the start door opened and stopped when the first floor switch, outside the start box, was hit.
2. *Total running time* was measured by a Standard Electric timer that started when the first floor switch was hit and stopped when the final floor switch was hit, closing the end-box door behind *S*. This provided the orthodox running measure, including all stops, reversals, and retracings as well as forward progress.
3. *Forward-progress* (FP) time was measured by a Standard Electric timer operated manually by *E*, who sat on a platform slightly above and to the side of the runway. When *S* stopped, *E* stopped this clock, which started when *S* hit the first switch. Whenever *S* reversed in the runway *E* also turned on the nearest marking light, to aid in remembering the position of farthest progress. When that position was again reached by *S* in his forward advance, *E* started the clock. The clock stopped when *S* hit the final floor switch. The FP time was thus a measure of the time *S* required to traverse the entire 14-ft. runway, without stopping, reversing, or retracing.
4. *Direct-run* (DR) time was the running time on those trials when no stops, and thus no gross competing behavior, occurred. All such trials were marked on the record sheet by *E*. The DR time was thus the purest measure of noncompetitive performance, and was also completely automatically obtained.

Extinction.—One trial per day was given for a maximum of 72 days. The only difference from acquisition procedure was that the burette was removed from the end box, and the 20 sec. were timed from entrance into the end box. Each *S* was continued to a criterion of two successive failures to reach the end box within 180 sec.

RESULTS

Acquisition

Typical changes in orthodox starting and total running time measures were obtained but are not reported in detail here. Starting time is largely a measure of competing activity, which decreases as *S* orients itself more effectively for the start-door opening. Competing activity in the runway is very slight at the 23-hr. deprivation level, as Cotton's (1953) data indicated. Our total running scores, including extraneous behaviors, were in all cases parallel and very close to the FP score (cf. Fig. 4).

The FP and DR scores were transformed from time scores to speeds, in terms of ft/sec, and all graphs and statistical analyses utilized this measure.

The FP speeds for the entire 73 trials are shown in Fig. 1, in blocks of six trials (seven for the final block). Here the concentration effect was not reliable, $F(3, 24) = 1.21$, $p > .05$. The blocks effect was highly reliable, $F(11, 264) = 26.17$, $p < .005$; and the Concentration × Block interaction was reliable, $F(33, 264) = 1.616$, $p < .05$.

An interesting feature of Fig. 1 is the decline from asymptote of each of the concentration groups and the subsequent return to a new and apparently lower asymptote. That this phenomenon is a genuine one is suggested by the fact that the decline occurred in each of the concentration groups and persisted over a relatively long period of time. Furthermore, it came at a progressively later point for the various concentrations, occur-

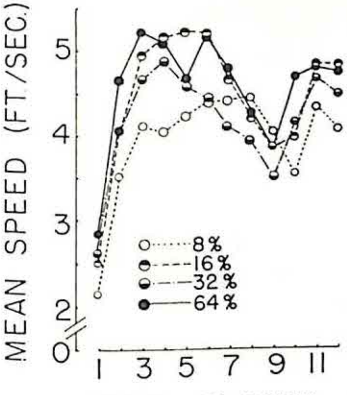


FIG. 1. Mean forward-progress speeds by sucrose concentrations for 12 blocks of six trials in acquisition.

ring first for the 64% *S*s, which hit their first asymptote earliest, next for the 32% *S*s, which hit their first asymptote next (albeit a lower one than that of the 16% *S*s), and then for the 16%, and finally for the 8% *S*s. Recovery from the decline in speed was also in that order.

A separate analysis of variance of the FP speeds by sucrose concentrations was performed for the first two trial blocks (12 trials), in order to evaluate concentration effects over the early acquisition trials. Here the concentration effect was reliable, $F(3, 24) = 3.77$, $p < .05$. The effect was due primarily to slower running by the 8% group. An orthogonal comparison of the mean for the 8% group with the other group means showed that this group ran reliably slower, $F(1, 24) = 8.37$, $p < .01$. The trials effect was highly reliable, $F(11, 264) = 250.14$, $p < .001$; the Concentration × Trials interaction was also reliable, $F(33,264) = 2.02$, $p < .05$.

The DRs began early in acquisition, from the first (2 *S*s) to the eighth (1 *S*) trial, with a median number of only two trials before the first such run. There was some tendency for higher-concentration *S*s to show more DRs early (group totals in the first six-trial block of 27, 18, 11, and 10 for 64%, 32%, 16%, and 8% groups, respectively), but by the sixth trial most of the *S*s were running directly, and this conditon continued throughout the 73 acquisition trials.

Mean DR speeds for the four concentration groups are shown in Fig. 2. Analysis of variance of these data showed reliable main effects for concentration, $F(3, 75) = 3.47$, $p < .025$; and trials, $F(11, 208) = 27.80$, $p < .001$. The increment over trials and the decline from the first asymptote are both quite clear. Scheffé's (1959, pp. 362–363) test was utilized to determine if the decline and subsequent recovery were statistically reliable. To test the decline, means were compared for the fifth and ninth trial blocks. To compute the F^* value, Scheffé's recommendation that the .10 level of confidence be adopted, to increase the power of the test, was followed. The contribution of the decline to the trial blocks F was reliable at the $p = .10$ level ($F^* = 18.77$; 15.18 required for .10, 20.13 for .05). A test of the recovery was made by comparing Trial Block 9

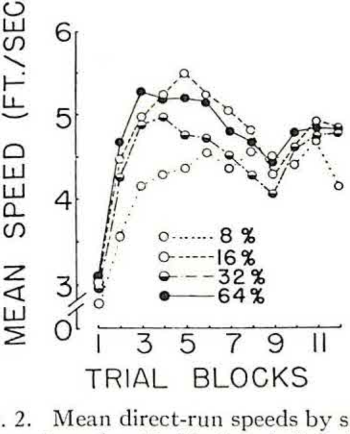


FIG. 2. Mean direct-run speeds by sucrose concentrations for 12 blocks of six trials in acquisition.

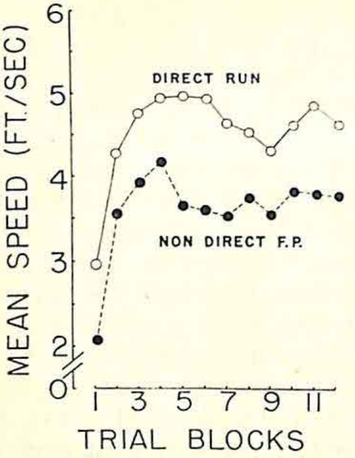


FIG. 3. Mean direct-run and nondirect-run forward-progress speeds in acquisition.

with Trial Block 11. This difference was not reliable at the .10 level ($F^* = 10.03$; 15.18 required).

When the four curves shown in Fig. 2 are combined into a single curve of DR speeds for all *S*s, and a comparable curve for all *S*s' FP speeds on *nondirect* runs is plotted on the same axes, the result is shown in Fig. 3. There is no question as to the reliability of the overall difference; as one might expect, every one of the 26 *S*s available for the comparison[4] ran faster on DR than on non-DR trials (mean difference of .88 ft/sec).

Extinction

No consistent group differences appeared in extinction in either the FP or DR speed measures, except for a slight tendency for the 8% *S*s to show a lower FP speed during the early trials. Sucrose concentration groups were therefore pooled, and compared with similarly pooled curves for acquisition. From the fourth block, extinction means were computed on the basis of a regularly diminishing *N* (to the final 13 *S*s which did not meet the criterion), but in spite of this selection the scores remained reasonably stable once the asymptotes were reached. Figure 4 shows this acquisition-extinction comparison, together with the corresponding curves for the total-running measure (total time on all trials included). FP speed is plotted because of the scarcity of DR trials after the first few days of extinction. Criterial extinction trials are excluded.

Two major aspects of Fig. 4 are to be noted. First, there was a marked reduction in FP speed in extinction, quite similar to the increment that occurred during acquisition. Note, nonetheless, that the asymptotic extinction speed by this measure was very close to the original acquisition level. Second, while the two running measures showed very similar progress during acquisition, consistent with Cotton's (1953) finding for rats under a relatively high-drive condition, they did diverge markedly during extinction, and the total-running speed

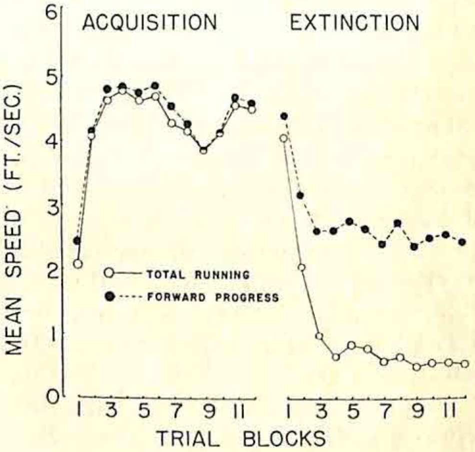


FIG. 4. Mean total-running and forward-progress speeds in acquisition and extinction.

---

[4] Two *S*s could not be used for this comparison because they did not have a sufficient number of non-DR runs (10 of the 12 blocks of six trials in each case being entirely composed of direct runs).

declined to a level markedly below the original level in acquisition.

To test the apparently greater divergence of these two measures in extinction the following procedure was adopted. For each *S* total-running speed during the first acquisition block was subtracted from total-running speed during the fourth block; the corresponding difference in speed between the first and fourth blocks was also obtained for the FP measure. The difference between these two measures of gain in speed in acquisition was then compared with the comparable difference in loss of speed over the first four trial blocks in extinction. The result of this final comparison was that 26 of the 28 *S*s showed a relatively greater change in extinction than in acquisition. The mean difference between scores was 1.46 ft/sec; this mean score was reliably greater than zero, $t(27)=8.95$, $p<.001$. It is therefore concluded that the differential loss in extinction was reliably greater than the differential gain in acquisition, as indicated in Fig. 4.

The number of DRs per rat in extinction varied from a minimum of 2 to a maximum of 11 with a median and a mode (7 *S*s) of 5. In order to obtain sufficient data for a statistical test of a change in DR speed over extinction some compromise had to be made. The 11 *S*s which had at least 7 DR scores in extinction were therefore selected for statistical analysis. Mean DR speeds over these scores were, from first to seventh, 5.22, 5.30, 5.13, 4.84, 5.03, 4.18, and 4.29 ft/sec. The decline in speed was reliable, $F(6, 60) = 20.18$, $p < .001$. Nine of the 11 *S*s showed a decrement from the first to the seventh DR score.

The same trend, and an even more pronounced loss in speed over the last few scores (to 3.67 and 3.57 ft/sec, for the final two DRs), is evident when all of the DR data are considered. Finally, only 3 of the 28 *S*s showed a faster speed on their last DR compared to their first DR in extinction. It is therefore concluded that no serious artifact from selection was involved in the preceding analysis and that the decline in DR speed over extinction was a genuine one, comparable to the increment over acquisition trials (Fig. 2).

Resistance to extinction as a function of sucrose concentration showed a slight curvilinear trend. The 8% *S*s were rather clearly most resistant, with 6 of 7 failing to extinguish within the 72 trials run. This number is to be compared with three, one, and three failures to extinguish in the 16, 32, and 64% groups, and the difference here was not quite reliable at the .05 level, $\chi^2(3) = 6.35$, $p_{.05} = 7.82$. Median runs to extinction were 72, 46, 22, and 42, in order of increasing sucrose percentage. The corresponding medians for trials to the first failure to run within 180 sec.[5] were 64, 19, 16, and 21. Three of the 8% *S*s ran on every one of the 72 daily extinction trials; only 1 of the *S*s in the three other groups did so. Comparison of the 8% and 16% means (50.0 compared with 22.9 trials) suggested the reliability of this difference, $t(12) = 2.32$, $p < .05$, although an analysis of variance of scores for all four groups indicated reliability slightly short of the .05 level, $F(3, 24) = 2.587$, $p_{.05} = 3.01$.

DISCUSSION

The results of this experiment offer a clear and straightforward positive answer to the primary empirical question under consideration: running behavior with gross competing responses eliminated shows a regularly increasing speed during acquisition and a regularly decreasing speed during extinction. The evidence is especially impressive when the DR measure is considered because, as mentioned earlier and suggested by Fig. 3, this measure is intrinsically clearer, in terms of competing responses, than the FP measure. The earlier runway results of King (1959), Cicala (1961), and Kintsch (1962), obtained under different experimental conditions, are supported.

---

[5] One *S* in the 32% group died after 49 trials and so could not be assigned a normal extinction score. However, this *S* was given a score for the less rigorous criterion of first failure to run.

There is some evidence in the present experiment that incentive variation can directly influence running behavior with competing responses controlled. However, this evidence is not as convincing as that for the practice and extinction effects. It mainly involves, within the present range of incentive values, the reduced speed evoked by the 8% concentration of sucrose.

It is of course possible that even as the rat continues to run down the runway his speed is influenced by competing responses which are not entirely incompatible with running—that is, he does not stop completely—but which do tend to slow him down. Such behaviors, for example, as looking, sniffing, and the like, might well occur and interfere with running speed differentially over trials. But such behaviors will need to be specified, and more analytic observational and recording techniques will need to be developed than have heretofore been used, before this alternative to a direct relationship between motivation and running vigor can be convincingly demonstrated.

The marked decline in running speed from early asymptote in each of the sucrose concentration groups is especially interesting in view of the clearly spaced nature of the reinforcement (cf. Prokasy, 1960). Interpretation in terms of satiation and cutoff factors (e.g., Collier & Willis, 1961) is suggested by the fact that the time of decline was perfectly ordered by sucrose percentage, with the highly concentrated 64% incentive producing the earliest effect. Although only 20 sec. of daily licking was permitted, involving a total liquid volume of less than 1 ml., apparently even this relatively small amount of sucrose provided daily is sufficient to produce some satiation. However, the relatively low degree of volume ingested and correspondingly small amount of ingestive activity suggest that the temporary satiation effect was related to a loss of incentive value per se, rather than mediated by the physiological factors known to operate over longer intake intervals.

The fact that the role of competing responses was considerably greater in extinction than in acquisition, as indicated in Fig. 4, is particularly relevant to a motivational interpretation of extinction (Marx, 1958, 1960; Marx & Murphy, 1961). The motivational theory is also supported by the fact that with competing responses controlled running speed not only showed a clear and progressive increment during acquisition but also showed a corresponding decrement during extinction.

Methodologically, the present data offer strong support to the assumption of the value of the more analytic speed measures, especially the DR score. Questions are raised as to the adequacy of the total-time measure in a number of standard problems where it has been widely and largely uncritically employed. For one example, the whole question of whether or not a gradient of running occurs (cf. Collier, Knarr, & Marx, 1961; Knarr & Collier, 1962) can be appropriately investigated only if segmental scores are clearly free of stops and retraces. A preponderance of such competitive and time-consuming behaviors toward the start of the runway early in trials, as commonly seems to occur under certain conditions (cf. Marx, 1950), could easily produce an artifactual "goal gradient"—that is, one not attributable to differences in running speed per se, as is ordinarily assumed.

REFERENCES

CICALA, G. A. Running speed in rats as a function of drive level and presence or absence of competing response trials. *J. exp. Psychol.*, 1961, **62**, 329–334.

COLLIER, G., KNARR, F. A., & MARX, M. H. Some relations between the intensive properties of the consummatory response and reinforcement. *J. exp. Psychol.*, 1961, **62**, 494–495.

COLLIER, G., & WILLIS, F. M. Deprivation and reinforcement. *J. exp. Psychol.*, 1961, **62**, 377–384.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

KING, R. A. The effects of training and motivation on the components of a learned instrumental response. Unpublished doctoral dissertation, Duke University, 1959.

KINTSCH, W. Runway performance as a function of drive strength and magnitude of reinforcement. *J. comp. physiol. Psychol.*, 1962, **55**, 882–887.

KNARR, F. A., & COLLIER, G. Taste and consummatory activity in amount and gradient of reinforcement functions. *J. exp. Psychol.*, 1962, **63**, 579–588.

MARX, M. H. Experimental analysis of the hoarding habit in the rat: I. Preliminary observations. *J. comp. physiol. Psychol.*, 1950, **43**, 295–308.

MARX, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

MARX, M. H. Resistance to extinction as a function of degree of reproduction of training conditions. *J. exp. Psychol.*, 1960, **59**, 337–342.

MARX, M. H., & MURPHY, W. W. Resistance to extinction as a function of the presentation of a motivating cue in the startbox. *J. comp. physiol. Psychol.*, 1961, **54**, 207–210.

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

PROKASY, W. F. Postasymptotic performance decrements during massed reinforcements. *Psychol. Bull.*, 1960, **57**, 237–247.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

(Received March 6, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 190–193

PAIRED-ASSOCIATE LEARNING AS A FUNCTION OF AROUSAL AND INTERPOLATED INTERVAL[1]

LEWIS J. KLEINSMITH AND STEPHEN KAPLAN

University of Michigan

This experiment tested the hypothesis that due to the phenomenon of perseverative consolidation, a pattern perceived under high arousal should show stronger permanent memory and weaker immediate memory than a pattern accompanied by low arousal. While recording skin resistance as a measure of arousal, 48 Ss were presented 8 paired associates for learning. The Ss were tested at various time intervals: 2 min., 20 min., 45 min., 1 day, and 1 wk. The results confirmed the hypothesis ($p = .001$). Paired associates learned under low arousal exhibited high immediate recall value and rapid forgetting. High arousal paired associates exhibited a marked reminiscence effect, that is, low immediate recall and high permanent memory.

There is growing evidence that perseverative consolidation of the memory trace can last over a considerable period of time. Glickman (1961) provides an extensive review of the literature in this area. Still more recently, Paré (1961), by the use of drugs, has shown that arousal plays an important role in the consolidation process. A stimulant administered immediately following a learning trial increases retention, while a depressant has the opposite effect.

As Paré has indicated, reverberating neural circuits (Hebb, 1958) provide a mechanism to explain the effect of arousal on consolidation. Under conditions of low arousal, relatively little nonspecific neural activity will be available to support the reverberating trace, resulting in little consolidation and poor long-term retention. On the other hand, under conditions of high arousal the increased nonspecific neural activity will result in more reverberation, and thus retention should be better.

While reverberation is taking place, however, one might expect the trace to be relatively unavailable to the organism, resulting in poor recall of the consolidating material during this interval. Such unavailability follows from a consideration of the difficulties in refiring a neuron which is already firing repeatedly in a reverberating circuit. Hodgkin (1948) has shown that neurons are sharply limited in their maximum rate of firing. Thus under high arousal, perseverative consolidation should be effective both in strengthening the memory trace and in making the memory difficult to evoke until the perseveration terminates.

In the present study *S*s were asked to learn a set of paired associates composed of words as stimuli and single-digit numbers as responses. The words were chosen to differ in their arousal value (as measured by galvanic skin response). In terms of the hypothesis, word-number pairs of high arousal value should be recalled poorly at first, but should be recalled well at a later time. Low arousal pairs, by contrast, should be remembered better at first and should show a gradual decay (forgetting) with time.

---

[1] This investigation was supported in part by a research grant (M4239) from the National Institutes of Health, Public Health Service. The *S*s in this study were run by Robert D. Tarte, research assistant on this project.

METHOD

Subjects.—The *S*s were 48 University of Michigan undergraduates obtained from introductory courses in psychology. They were run in six subgroups of 8 *S*s each.

Procedure.—The *S*s were given a single learning trial with a list of eight word-number pairs. Eight words expected to produce different arousal levels were chosen as stimulus words. These were KISS, RAPE, VOMIT, EXAM, DANCE, MONEY, LOVE, SWIM. The response items were single digits from 2 to 9.

A slide projector with an exposure time of 4 sec. was used to present the stimuli. During the training trial *S* first saw the stimulus word alone, and then saw the word repeated with a single digit response. To separate the arousal effects from one set of pairings to the next, two slides containing four colors each were inserted between the paired associates, and *S* was instructed to name the colors. (Red, green, blue, black, yellow, and orange were used randomly on these slides.) The *S* was instructed to "concentrate carefully on both colors and word-number pairs" as he called them out loud, but to avoid rehearsal *S* was not specifically told that he would be tested for recall.

In order to determine the arousal effects of each stimulus word, skin resistance was recorded during learning. The electrodes were of the zinc variety described by Lykken (1959). To insure constancy of conditions, electrodes and recording apparatus were also used during the recall session.

During the recall session *S* was instructed to indicate the correct number for each stimulus word as it appeared. The correct numbers were not repeated. Colors were used as an interpolated task as before.

Design.—The six groups were tested one at each of the following recall intervals: immediate (about 2 min.), 20 min., 45 min., 1 day, and 1 wk. Two groups were run at the first recall interval.

To correct for serial order effects eight different training lists were used for each group, designed so that each of the eight words appeared in each position in the list once (Fisher & Yates, 1938). The order of the recall lists was varied in the same manner.

Data analysis.—Any drop in *S*'s skin resistance which occurred within 4 sec. of presentation of a given word was considered an arousal deflection; these were converted to percent deflections. Each *S*'s eight GSR deflections were then ranked. (In case of ties, absolute skin resistance level was used to break them. A deflection occurring at a low level of absolute skin resistance was considered higher arousal than a similar deflection occurring at a higher absolute level.) The three highest deflections for each *S* were designated "high arousal learning," and the three lowest were designated "low arousal learning."

RESULTS

Figure 1 illustrates the relationship between high and low arousal recall as a function of time. At immediate recall, numbers associated with low arousal words are recalled five times as often as numbers associated with high arousal words. The capacity to recall numbers associated with low arousal words decreases as a function of time in a normal forgetting curve pattern. On the other hand, the capacity to recall numbers associated with high arousal words shows a considerable reminiscence effect. After 20 min., the increase is more than 100%, and after 45 min. it has increased 400%. This high capacity for recall of high arousal pairs persists for at least a week—the longest interval employed.

Analysis of variance (Lindquist, 1953) for correct responses confirms the trends in the figure at $p < .001$

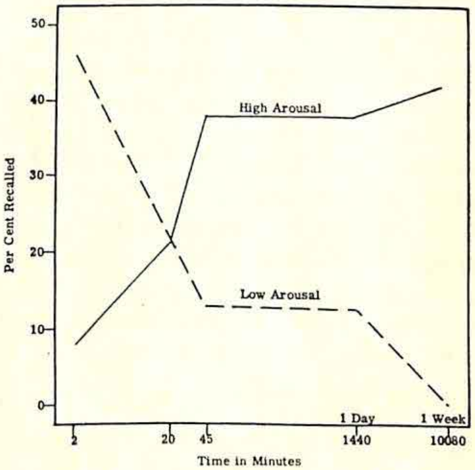


FIG. 1. Differential recall of paired associates as a function of arousal level.

TABLE 1

Analysis of Variance of Recall Scores

Source	*df*	*MS*	*F*
Between *S*s	47		
Conditions (B)	4	0.1459	.27
Error (b)	43	0.5377	
Within *S*s	48		
Arousal level (A)	1	0.1667	.27
A × B	4	5.7396	9.45***
Error (w)	43	0.6076	
Total	95		

*** $p < .001$.

(Table 1). The significant interaction is primarily attributable to the effect of the immediate condition, $F(1, 15) = 11.36$, $p < .005$, and the 1-wk. condition, $F(1,7) = 11.67$, $p < .025$. Thus at immediate recall, low arousal learning is significantly greater than high arousal learning; after 1 wk., the situation has reversed and high arousal recall is significantly better than low arousal recall.

As can be seen in Table 2 the distribution of items at different times sampled forms no systematic patterns that might account for the results obtained. This table shows the proportion of each of the eight items at each time sampled for both high and low arousal learning.

Discussion

The increase in the capacity to recall items learned under conditions of high arousal in contrast to items learned under low arousal conditions provides further support for the theory of reverberating neural circuits. A somewhat simplified physiological explanation of the processes involved may be pictured as follows. When a person perceives a pattern, a closed, reverberating neural circuit is set up in his brain corresponding to this pattern. The more nonspecific neural activity or arousal present, the greater the number of times which the trace is likely to reverberate. And the greater this perseverative consolidation of the neural trace, the stronger the permanent memory.

However the apparent paradox is that while perseveration is taking place, recall ability is poor. This follows from the hypothesis that at any given instant the neurons involved in the perseverating trace are either already in the process of firing, or are in an absolute refractory state, or may be in a state of slowly developing subthreshold activity (Hodgkin, 1948); thus the trace would be relatively unavailable to the organism. With greater arousal there will be increased perseveration and thus poorer immediate performance.

This explanation coincides with the empirical findings. High arousal learning showed poor immediate recall due to

TABLE 2

Distribution of Items at Different Times Sampled

Item	High Arousal					Low Arousal				
	2 Min.	20 Min.	45 Min.	Day	Wk.	2 Min.	20 Min.	45 Min.	Day	Wk.
DANCE	.06	.08	.13	.08	.13	.13	.08	.21	.21	.17
EXAM	.13	.04	.17	.08	.13	.15	.13	.08	.17	.21
KISS	.10	.13	.08	.17	.08	.15	.08	.08	.04	.17
LOVE	.10	.13	.13	.13	.17	.17	.17	.13	.17	.08
MONEY	.10	.17	.08	.04	.08	.08	.17	.17	.21	.13
RAPE	.27	.13	.21	.21	.25	.04	.17	.04	.08	.04
SWIM	.10	.13	.13	.08	.08	.15	.13	.13	.08	.13
VOMIT	.13	.21	.08	.21	.08	.15	.08	.17	.04	.08

the hypothesized period of unavailability of the trace, while recall was better under low arousal, for this immediate period. The long-term effects are, as predicted, in the opposite direction. There was significantly poorer long-term retention under low arousal and superior long-term retention under high arousal. This is in keeping with the expectation that there would be more consolidation and thus better learning under high arousal.

Reminiscence has become so rare in recent years that Underwood (1953) has wondered if it might not be a prewar phenomenon. This study suggests that methodological considerations might be an important factor in the return of reminiscence. The innovation in this experiment consisted of measuring arousal during learning and thus independently determining in advance which items should show reminiscence effects and which should not. When recall as a function of time is plotted for all items at once, little reminiscence is evident.

REFERENCES

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver and Boyd, 1938.

GLICKMAN, S. E. Perseverative neural processes and consolidation of the memory trace. *Psychol. Bull.*, 1961, **58,** 218–233.

HEBB, D. O. *A textbook of psychology.* Philadelphia: Saunders, 1958.

HODGKIN, A. L. The local electric changes associated with repetitive action in a nonmedullated axon. *J. Physiol.*, 1948, **107,** 165–181.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

LYKKEN, D. T. Properties of electrodes used in electrodermal measurement. *J. comp. physiol. Psychol.*, 1959, **52,** 629–644.

PARÉ, W. The effect of caffeine and seconal on a visual discrimination task. *J. comp. physiol. Psychol.*, 1961, **54,** 506–509.

UNDERWOOD, B. J. Learning. *Annu. Rev. Psychol.*, 1953, **4,** 31–58.

(Received March 23, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 194–200

SEMANTIC TRANSFER OF THE DIFFERENTIAL CONDITIONED EYELID RESPONSE FROM WORDS TO OBJECTS[1]

THOMAS F. HARTMAN[2]

University of Wisconsin

This experiment investigated transfer of the eyelid CR from words to objects possessing characteristics of the words and evaluated the effects of concurrently verbalizing specific characteristics of the objects. 120 Ss were used. Findings were (a) marked discrimination early in the transfer phase even when the words initially differentially reinforced were unrelated to the objects; (b) more initial CR transfer to the positive object CS for 2 groups with initially relevant words—one group verbalizing a dimension of the objects identical to the words, the other group verbalizing another relevant dimension of the objects; (c) more initial transfer of discrimination only for the group verbalizing the dimension of the objects identical to the word CSs.

The purpose of the present investigation was to explore the transfer of the eyelid CR from word CSs to object CSs possessing characteristics denoted by the words, and to evaluate the effects on transfer of verbalizing a relevant characteristic of the object CS that was or was not represented by the initial words CSs. The few conditioning studies on word-object transfer in available literature (Branca, 1957; Kapustnik, 1930) as well as those investigating other types of semantic transfer (See Razran, 1961) have used relatively long latency CRs (e.g., GSR, salivation, vasoconstriction, blood coagulation). It seems desirable to explore semantic transfer with a shorter latency response, as shortening time for response should reduce variability in behavior, and may also limit the amount of semantic mediation that can occur. Since the methodology of the eyelid CR has been more thoroughly explored than the CRs previously used (Grant, 1945; Grant & Norris, 1947; Hartman, Grant, & Ross, 1960; Hartman & Ross, 1961; Spence & Ross, 1959; Spence & Taylor, 1951), study of semantic transfer of this response would be desirable in any event. One methodological complication of the eyelid CR offers an unusual opportunity to investigate the contribution of voluntary responding. Because *S*s in eyelid conditioning can be separated into voluntary and nonvoluntary responders solely on the basis of the form of the response (Hartman & Ross, 1961; Spence & Ross, 1959) it is possible to determine objectively the differences between these two modes of responding in semantic transfer of differential conditioning.

METHOD

Apparatus.—The apparatus was modified from that used in earlier research (Hartman & Grant, 1960). Ambient room illumination was approximately 1 mL. Immediately in front of *S* was a microphone which was connected to a noise operated relay. About 100 cm. in front of *S*'s eyes and approximately at his line of level vision was a rectangular ground glass screen 45 cm. wide and 30 cm. high. During the first part of the experiment the CS consisted of flashing a word on the screen with a 16-mm. strip film projector. The words were in block letters, white on a dark background, approximately 6 cm. high. Between presentations of words the screen displayed a frame of the same density as the background of the frames containing words. On each side of the screen at the level of the projected words was a 10-cm. circular milk glass window. The centers of the two windows were 62 cm. apart horizontally. The illumination of either of these windows was the CS in the second part of the experiment. The window on the left was illuminated by a blue 6-w. lamp; the window on the right by a similar red lamp. Neutral filters were inserted in the windows to equalize their apparent brightness at about .8 mL. The final hues were desaturated blue and pink which were not evident unless the windows were illuminated. The CS in both parts of the experiment had a duration of 1250 msec. On reinforced trials, the UCS was a puff of nitrogen of 50 msec. duration administered to *S*'s cornea 1060 msec. after the onset of CS. The puff was monitored by a 150-mm. column of mercury at the source. Intertrial intervals varied between 15 and 35 sec. with a mean of 25 sec. as programed by a tape transmitter.

---

[1] Based upon a dissertation submitted to the Psychology Department, University of Wisconsin in partial fulfillment of the requirements for the PhD degree. The author wishes to thank David A. Grant for his assistance. This research was supported in part by the National Science Foundation and in part by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation.

[2] Now at Thomas J. Watson Research Center, IBM, Yorktown Heights, New York.

Experimental design.—Before the session began *S* was read the "neutral" instructions used at the Wisconsin laboratory. A 3 × 2 orthogonal design was used. In the first part of the session three groups of 40 *S*s each were used; one received the words LION and DEER as CSs, a second the words PINK and BLUE, and a third the words LEFT and RIGHT. The words were presented with one word of the pair consistently reinforced for half the *S*s, and the other word of the pair consistently reinforced for the other half of the *S*s. All *S*s were given 20 reinforced trials and 20 unreinforced trials with each successive block of 10 trials containing 5 reinforced and 5 unreinforced trials. The sequences were further constrained so that runs of like trials greater than 3 did not occur. One random sequence and its complement were used in each group. Each of the sequences was used with each word of the pair reinforced, so there were four complete word-reinforcement sequences in each group.

During the minute between the two parts of the experiment half the *S*s in each group were told to report the color and half the *S*s were told to report the laterality of the light that appeared as the CS. Either the window on the left (blue CS) or the window on the right (pink CS) was consistently reinforced; the reinforcement of laterality or color was consistent with the words that were reinforced in the groups that received the words LEFT and RIGHT or PINK and BLUE as the CSs. All *S*s were given 20 reinforced trials and 20 unreinforced trials to the lights, incorporating the same restraints on randomness as in the first part of the experiment.

In referring to the groups in the present study, the CS words are typed in all capital letters followed by the overt verbalizations in both capital and lower case letters. For example, the group initially conditioned to the word pair LION and DEER that verbalized the color of the objects in the transfer stage will be designated as LION-DEER/Pink-Blue.

Subjects.—The *S*s were 72 men and 48 women students at the University of Wisconsin. Each of the three groups that received different word pairs contained equal numbers of men and women. Assignment to these groups was random within sex, but an attempt was made to equate the frequency of response to the reinforced word during the first stage of the experiment for the two groups that were later to make different verbalizations to the CS lights within each word-pair group.

RESULTS

All eyelid closures with latencies from 200 msec. to 10 msec. before *S*'s reflex wink to the UCS were considered anticipatory responses. In some analyses, specified below, *S*s giving voluntary responses were removed from the data, using the criterion suggested by Hartman and Ross (1961); i.e., *S*s with time derivatives of anticipatory eyelid responses greater than 35% of the mean derivative of the reflex to UCS on more than half their responses were classified as voluntary *S*s. Since the data of greatest interest in the present study were the trials immediately after the shift from word CSs to object CSs, the criterion for rejection of voluntary *S*s was applied only to the first five postshift trials to the positive CS.

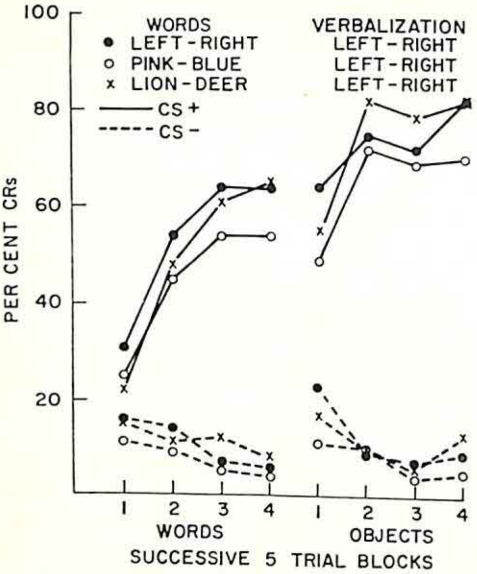


FIG. 1. Percentages of CRs plotted for successive five-trial blocks for the groups verbalizing the position of the objects. (Plots are from the data of all *S*s.)

Figures 1 and 2 show the mean percentage of anticipatory responses to the positive and negative CSs for all *S*s in each group plotted as a function of successive five-trial blocks. When *S*s giving voluntary form responses were removed the acquisition functions were basically the same except that the level of response to the positive CSs was somewhat lower; therefore, they are not presented. The derivative criterion classified 20 *S*s as voluntary: 2 in the PINK-BLUE/Left-Right group, 3 each in the LEFT-RIGHT/Left-Right and LION-DEER/Pink-Blue groups, and 4 each in the remaining groups. Analyses of variance among reinforcement patterns for each word pair and among reinforcement patterns for the objects were not significant, so the pattern of reinforcement was ignored in these figures and all computations.

Conditioning to word pairs.—It is apparent from Fig. 1 and 2 that marked conditioned discrimination can be established to word CSs. The terminal conditioned discrimination was significant in every group; analyses of variance of frequency of response over the last five-trial blocks gave highly significant *F*s between positive and negative CSs; the smallest $F(1, 18) = 19.66$ for the data of all *S*s in the PINK-BLUE/Left-Right group. To determine if any of the words produced different terminal response rates, separate analyses of variance of frequency of response among word pairs over the last five-trial blocks were computed for all *S*s and nonvoluntary *S*s for both the positive and negative CSs. The largest F among the three word pairs was $F(2, 97) = 1.46$ ($p > .25$). All words used were, therefore, approximately equally effective when used as CSs.

Initial transfer to objects.—Figures 1 and 2 show considerable discrimination to the objects during the first five-trial transfer blocks even if the word pair originally reinforced was not related to the objects. The condi-

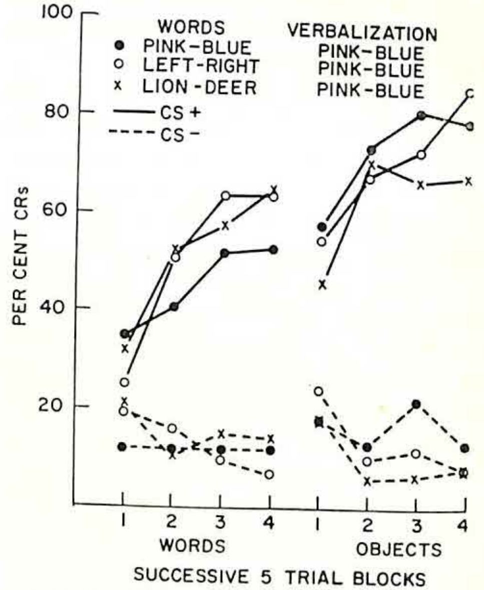


FIG. 2. Percentages of CRs plotted for successive five-trial blocks for the groups verbalizing the color of the objects. (Plots are from the data of all *S*s.)

tioned discrimination over the first five-trial blocks was tested in each group by analysis of variance. The discrimination was highly significant in every group, the smallest F between positive and negative CSs was $F(1, 18) = 22.98$ in the LION-DEER/Pink-Blue group. The curves for the positive CSs show consistent differential transfer of frequency of response in both figures; i.e., the groups with the initial word pair relevant to the objects and the concurrent verbalization a continuation of the initial word pair showed the most transfer, and the groups with the initial words irrelevant to the objects showed the least transfer. The curves for the negative CSs show no consistent differential transfer of frequency of response. After the first block of trials the stimuli produced essentially equivalent rates of response regardless of the initial CS words.

In order to have the most sensitive measure of differential transfer a dependent variable reflecting magnitude as well as frequency of response was used. The measure incorporating response magnitude was the ratio of the amplitude of the CR to the amplitude of the reflex response to the UCS on that trial for a reinforced trial, and for an unreinforced trial the ratio of the amplitude of the CR to the amplitude of the reflex response to the UCS on the preceding reinforced trial for the word CSs or the following reinforced trial for the object CSs. If this ratio on an unreinforced trial was greater than 1, it was arbitrarily recorded as 1. The percentage transfer of CRs to the positive CS summarized in Table 1 illustrates that although this magnitude ratio score (MR) shows the same trend of differential transfer as does percentage frequency, it is sensitive to the fact that many of the CRs to the objects following an irrelevant word pair are quite small. Table 1 shows this was most evident for the nonvoluntary *S*s; voluntary *S*s tend to give more complete closures on both original stage and transfer stage. In this study nonvoluntary *S*s gave very few complete closures—only 13% of the responses over the last 10 trials to the word CSs.

TABLE 1

PERCENTAGE TRANSFER OF CRs TO THE POSITIVE CS FROM THE LAST FIVE PRESHIFT TRIALS TO THE FIRST FIVE POSTSHIFT TRIALS

Group	All *S*s		Nonvoluntary *S*s	
	Frequency	MR	Frequency	MR
LEFT-RIGHT/Left-Right	100.0	98.8	103.9	101.4
PINK-BLUE/Left-Right	90.7	91.3	93.3	94.0
LION-DEER/Left-Right	84.6	78.2	86.8	79.1
PINK-BLUE/Pink-Blue	109.4	105.0	118.3	132.8
LEFT-RIGHT/Pink-Blue	85.9	88.4	89.0	92.0
LION-DEER/Pink-Blue	70.8	64.3	67.3	59.2

Differential transfer was expected to be most evident in the first few trials to the objects, so that to test for differential transfer only the data for the first five-trial blocks to the objects were used. Separate analyses of covariance were computed for the data of all *S*s and the data of the nonvoluntary *S*s. The dependent variable was MR on the first five-trial blocks to the objects; the concomitant variable was MR on the last five-trial blocks to the words. The data were analyzed as a 2×3 orthogonal design, with the verbalizations Left-Right and Pink-Blue orthogonal to verbal relevance; i.e., (*a*) the initial word pair relevant to the objects and the verbalizations a continuation of the word pair, (*b*) the initial word pair relevant to the objects but the verbalizations another dimension of the objects, or (*c*) the initial word pair not relevant to the objects. These analyses of covariance are summarized in Table 2. Whether *S* verbalized

TABLE 2

Analyses of Covariance of Summed MR on First Five Postshift Trials

Source	*df*	Positive CS	Negative CS	Difference[a]
		F	*F*	*F*
All *S*s				
Verbalization (V)	1	0.568	1.090	1.612
Verbal relevance (R)	2	7.737**	0.025	4.710*
V × R	2	0.829	2.090	1.451
Error (*MS*)	113	(5.431)	(4.219)	(6.394)
Among slopes	5	1.397	2.109	1.805
Slope error (*MS*)	108	(5.249)	(4.021)	(6.174)
Nonvoluntary *S*s				
Verbalization (V)	1	0.442	1.423	0.926
Verbal relevance (R)	2	8.260**	0.547	3.576*
V × R	2	1.246	2.806	0.648
Error (*MS*)	93	(5.689)	(3.461)	(6.097)
Among slopes	5	1.690	2.353	0.994
Slope error (*MS*)	88	(5.486)	(3.227)	(6.099)

[a] Difference = positive CS minus negative CS.
* $p = .05$.
** $p = .01$.

Left-Right or Pink-Blue was not significant in any of the analyses. There was, however, statistically significant differential transfer of responsiveness to the positive CS and statistically significant differential transfer of degree of discrimination with respect to verbal relevance for the analyses utilizing the data of all *S*s and the data of the nonvoluntary *S*s only. The *F*s were significant at the .01 level for responsiveness to the positive CS and the .05 level for degree of discrimination. The interaction was never significant; i.e., the same trends occurred in the groups that verbalized Left-Right as in the groups that verbalized Pink-Blue. Analyses of responses to the negative CS disclosed no significant effects.

The means adjusted by covariance that produced the significant *F*s in Table 2 are shown in Table 3. The means were examined using the multiple range test of Duncan (1955), with the technique of Kramer (1956) for unequal replication. Means not underlined by the same line are significantly different at the .05 level. Both groups with the initial word pair relevant to the objects showed significantly more transfer of responsiveness to the positive CS than the group with the initial words irrelevant, but the groups with the words relevant did not differ from each other. Only the group with the word pair relevant and the verbalization a continuation of the word pair showed significantly more transfer of discrimination than the group with the words irrelevant. For discrimination no other means were significantly different.

Latency of conditioned responses.—To determine if the mean latency of CRs was significantly related to the degree of discrimination, mean latency of CRs was correlated with the percentage discrimination in each of the treatment groups. The percentage frequency measure (number of responses to the positive CS divided by the total number of responses to the objects) was used rather than MR as it is a more common measure of discrimination in the existing literature. The correlations were computed from the data of the nonvoluntary *S*s, deleting

TABLE 3

Duncan Range Tests of Adjusted Summed MR to the First Five Postshift Trials

Response Measured	Irrelevant Initial Words	Relevant Words	Continuing Words
All *S*s			
Positive CS	1.71	2.09	2.35
Difference[a]	1.26	1.54	1.81
Nonvoluntary *S*s			
Positive CS	1.33	1.80	2.06
Difference[a]	1.07	1.26	1.56

Note.—Means not underscored by same line differ from each other, $p = .05$.
[a] Difference = positive CS minus negative CS.

3 additional *S*s who showed no conditioning (therefore no CR latency). The *r* average (*z* transform) over groups was .56 ($df = 79$), $p < .001$; *S*s with longer mean CR latency showed better discrimination. Analysis of mean CR latency disclosed no significant differences among the six groups.

Latency of verbalizations.—A test for relationship between latency of verbalization to the objects and percentage discrimination was made by correlating these data for the same 97 *S*s used in the CR latency analysis. No *S* failed to make the required verbalization on more than 1 of the 40 trials, but some of the verbalizations occurred at very long latencies. Any latency longer than 1235 msec. was arbitrarily recorded at that value, so that a few very long latency verbalizations did not markedly affect *S*'s mean latency of verbalization. The mean latency of verbalization was not significantly correlated with percentage discrimination; the *r* average over all groups was −.10 ($df = 79$), $p > .20$, indicating no linear relationship. Analysis of mean latency of verbalization disclosed no significant differences among the six groups.

DISCUSSION

The conditioned discrimination established to words in the present study was comparable to that reported by Hartman and Grant (1962) to laterally separated lights when a 1000-msec. CS-UCS interval was used. It appears that words as CSs gave approximately the same level of conditioning and discrimination as the laterally or horizontally separated lights typically used as the CSs in differential eyelid conditioning studies.

The formation of a differential conditioned response to the words produced marked facilitation of conditioned discrimination to the objects, as highly significant discrimination was evident in all groups at the end of the first five-trial blocks to the objects. This facilitation or transfer of discrimination was initially somewhat more evident in the groups conditioned to a word pair relevant to the objects, so there was a differential transfer attributable to the particular initial word pair in addition to the general facilitation from the formation of a previous discrimination. The large amount of transfer not dependent upon the particular word pair was noteworthy but not of primary interest in this investigation so that no direct statistical test for it was possible in this experiment; the design provided only for a test of differential transfer due to the relevance of the particular word pair to the verbalization concurrent with the presentation of the objects. It was possible, however, to determine indirectly the approximate magnitude of the general transfer of discrimination by comparing the discrimination on the first five-trial blocks to the objects with the discrimination on the first five-trial blocks to the word pairs. The discrimination to the objects was appreciably better. The discrimination on the first five-trial blocks to the objects was also much superior to that reported by Hartman and Grant (1962) on the first four-trial blocks to laterally separated lights with a 1000-msec. CS-UCS interval. Even without a direct statistical test the evidence for a general facilitative effect of a discrimination on a subsequent discrimination was quite convincing.

Despite the extensive general transfer of discrimination in all groups, there was consistent differential transfer among the groups for both responsiveness to the positive CS and degree of discrimination. The differential transfer of degree of discrimination was due entirely to the differential transfer of responsiveness to the positive CS, as there were no consistent trends in responsiveness to the negative CS among the groups. More transfer to objects when *S*s were previously conditioned to words relevant to the objects was expected, and the findings of Kapustnik (1930) and Branca (1957) were confirmed. Previous studies, however, provided no basis for predicting differences between the groups with relevant initial word pairs but different verbalizations to the objects. Stable semantic transfer with a skeletal CR does exist, but the effects were not large. Removing the *S*s from the data that gave voluntary form responses did not do away with the semantic transfer; it was not caused by *S*s responding voluntarily to the objects.

Apparently the verbalization need not be a continuation of the initial relevant word pair to obtain increased transfer; the data provided no explicit support for mediated transfer based on subvocal response to the initial words. The verbalizations in the present study only forced *S* to attend to a particular dimension of the objects; the latency of these responses was probably not a simple function of the discriminative process as no explicit set to vocalize as rapidly as possible was given. The *S*s did not adopt a "reaction time" set either, as the vocalizations of the *S*s in this study were considerably longer than the reaction times reported by Brainard, Irby, Fitts, and Alluisi (1962) under similar conditions. All nonvoluntary *S*s had a mean CR latency considerably longer than the estimate of the latency of the discriminative process from the data of Brainard et al. (1962), but only 45 of these 97 *S*s had a mean CR latency longer than the mean latency of verbalization. These data suggest that the overt verbalizations probably did not form a part of the effective stimulus complex that *S* was responding to in this study, but rather that the verbalizations were simply a resultant of the discriminative process, as were the CRs. Since overt verbalizations did not appear to be a part of the effective stimulus complex interpretation of the data in terms of Wickens' (Wickens, 1959; Wickens, Gehman, & Sullivan, 1959) mediated response model is not warranted.

REFERENCES

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., & ALLUISI, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

BRANCA, A. A. Semantic generalization at the level of the conditioning experiment. *Amer. J. Psychol.*, 1957, **70**, 541–549.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

GRANT, D. A. A sensitized eyelid reaction related to the conditioned eyelid response. *J. exp. Psychol.*, 1945, **35**, 393–402.

GRANT, D. A., & NORRIS, E. B. Eyelid conditioning as influenced by the presence of sensitized beta responses. *J. exp. Psychol.*, 1947, **37**, 423–438.

HARTMAN, T. F., & GRANT, D. A. Effect of intermittent reinforcement on acquisition, extinction, and spontaneous recovery of the conditioned eyelid response. *J. exp. Psychol.*, 1960, **60**, 89–96.

HARTMAN, T. F., & GRANT, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

HARTMAN, T. F., GRANT, D. A., & ROSS, L. E. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychol. Rep.*, 1960, **7**, 305–311.

HARTMAN, T. F., & ROSS, L. E. An alternative criterion for the elimination of "voluntary" responses in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 334–338.

KAPUSTNIK, O. P. [The interrelation between direct conditioned stimuli and their verbal symbols.] *Trudy Lab. Fiziol. Vyssh. Nervn. Deyatel. Rebyonka*, 1930, **2**, 11–22. (*Psychol. Abstr.*, 8:152)

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 1956, **12**, 307–310.

RAZRAN, G. The observable unconscious and the inferable conscious in current Soviet psychophysiology: Introceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–385.

SPENCE, K. W., & TAYLOR, J. Anxiety and the strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

WICKENS, D. D. Conditioning to complex stimuli. *Amer. Psychologist*, 1959, **14**, 180–188.

WICKENS, D. D., GEHMAN, R. S., & SULLIVAN, S. N. The effect of differential onset time on the conditioned response strength to elements of a stimulus complex. *J. exp. Psychol.*, 1959, **58**, 85–93.

(Received April 4, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 201–205

R-S LEARNING AND NEGATIVE TRANSFER EFFECTS WITH A MIXED LIST[1]

DONALD A. KAUSLER AND GEORGE A. KANOTI

St. Louis University

Transfer effects (both S-R and R-S) were investigated within a mixed list (C-D, C-B, A-C, A-B_r). 60 Ss learned List 1 (A-B) to a criterion and then received 10 trials on the mixed list (List 2). An R-S trial was given for each list. C-B pairs yielded significant negative transfer, confirming Twedt and Underwood's (1959) earlier finding. The amount of transfer for C-B pairs correlated with R-S score on C-B pairs ($r = .35$ for relative transfer), providing support for a confusion hypothesis. The paradigms differed significantly in R-S score on List 2, with C-D the greatest amount, then A-C, C-B, and A-B_r. C-D increased significantly in R-S recall from List 1 to 2, A-B_r decreased significantly, and A-C and C-B were unaffected.

The presence of R-S learning has been inferred in several transfer studies (Harcum, 1953; Murdock, 1956, 1958) for the purpose of predicting both positive and negative transfer in paradigms that deviate from the classical Bruce-Wylie paradigms. Moreover, the general importance of R-S learning in transfer effects within the Bruce-Wylie paradigms is suggested in a recent study by Twedt and Underwood (1959). They found an unexpected trend for negative transfer to occur in the traditional A-B, C-B paradigm under both mixed and unmixed list conditions. As a possible explanation they reasoned that the transfer may have resulted from a confusion between B-C (R-S) associations of the second list and B-A (R-S) associations of the first list.

The present study was concerned further with the relationship between R-S learning and negative transfer as related to both S-R learning of the second list and R-S recall on the second list. The study consisted largely of a replication of the mixed list part of the Twedt and Underwood (1959) experiment with the addition of R-S measures. The *S*s thus learned A-B pairs in List 1 and C-B, A-C, A-B_r, and C-D pairs in List 2, as in the original study. Three problems were investigated in the present study. The first concerned the generalization of the negative transfer found by Twedt and Underwood in the C-B paradigm to a different sample of *S*s. The second problem concerned the relationship between degree of R-S learning for the C-B pairs in List 2 and the amount of negative transfer for the C-B paradigm. This problem constituted a test of the hypothesis suggested by Twedt and Underwood to account for the existence of negative transfer in what is generally regarded as a positive transfer paradigm. The third problem involved the relative amounts of R-S recall on List 2 for each of the four paradigms. If R-S learning is related to transfer effects in List 2, then it might be expected that the effects would be reflected in the amounts of R-S recall per se.

---

[1] This study was supported in part by a grant from the National Science Foundation. The article is based on a thesis submitted by the junior author to the Graduate School, St. Louis University in partial fulfillment for the Master of Science degree.

METHOD

Lists.—The second list (List 2) was identical for all *S*s. It contained the same 12 pairs of two-syllable adjectives (low similarity) employed by Twedt and Underwood (1959) in their List 2. The first list (List 1) also contained 12 pairs of two-syllable adjectives (low similarity), with the C-D, C-B, A-C, and A-B$_r$ paradigms represented by three pairs each. That is, three pairs in List 1 were completely independent of List 2 content and, in conjunction with their List 2 counterparts, made up the A-B, C-D paradigm. Three other pairs in List 1 contained identical responses to their List 2 counterparts but completely different stimuli, thus making up the A-B, C-B paradigm. In a similar manner, the A-C and A-B$_r$ paradigms were represented by the appropriate List 1-List 2 relationships.

As in the Twedt and Underwood (1959) study, four conditions (I, II, III, and IV) were used in constructing List 1, with each condition represented by three pairs from List 2. The use of four groups of *S*s, one group for each condition, made it possible to use all 12 pairs in List 2 equally often within the various paradigms. These conditions controlled for the possibility of differential difficulty of items within the paradigms and provided a broader sampling of items in each paradigm. New stimuli for the C-B paradigm and new responses for the A-C paradigm, as well as new pairs for the C-D paradigm, were provided by 12 pairs of adjectives selected from the lower end of the Melton and Safier calibration (Stevens, 1951).

Subjects.—The *S*s, 55 men and 20 women from general psychology classes at St. Louis University, were assigned alternately to five groups: Groups I, II, III, IV, and C, with 15 *S*s in each group. Groups I–IV represented the four list conditions described above and for purposes of determining transfer effects were combined into one large experimental group of 60 *S*s. Group C was intended to be a special control group performing on lists identical to one of the list conditions for the 60 experimental *S*s. Condition II was selected at random for this purpose.

Procedure.—Each *S* was given standard instructions regarding paired-associate learning and the anticipation method. This was followed by two trials on a warm-up list containing three pairs of adjectives comparable to those in the main lists. The transfer lists were presented on a Lafayette drum at a 2:2-sec. rate, with a 4-sec. intertrial interval. All *S*s practiced on List 1 until they reached a criterion of one perfect trial, and all *S*s received 10 trials on List 2.

For the 60 experimental *S*s, practice on List 1 was followed by a 2-min. free recall period during which they were asked to write down the stimuli from List 1. This recall period, in turn, was followed immediately by an R-S recall trial in which the responses from List 1 were presented on the drum in a randomized order, and without the stimuli, at a 2-sec. rate and 2-sec. interitem interval. On this trial the *S*s were asked to name the stimulus that had been paired with the given response. Practice on List 2 was followed by a similar free recall period and an R-S trial involving List 2 responses.

For the 15 control *S*s (Group C), List 1 performance was followed by a period of informal conversation for approximately the time (3 min.) required by the experimental *S*s to complete the free recall and R-S tasks. This group was included in the study to serve as a rough check on the possible effects that drawing *S*'s attention to R-S learning might have on both S-R and R-S List 2 transfer. The free recall and R-S tasks for List 2 were administered to this group following practice on List 2.

For both the experimental and control *S*s a 1-min. rest break was given prior to practice on List 2. This break followed the R-S trial on List 1 for the experimental *S*s and the conversation period for the control *S*s.

RESULTS AND DISCUSSION

S-R Transfer Effects

Transfer on List 2 was measured absolutely by the formula $E - C$ and relatively by $E - C/E + C$ (Murdock, 1957), where E is the number of correct responses made in the 10 trials for the three pairs in the paradigm and C the number of correct responses for the three pairs in the C-D paradigm. The transfer effects for the C-B, A-C, and A-Br paradigms are summarized in Table 1. All three paradigms yielded significant negative transfer relative to the null value of zero transfer as measured absolutely or relatively[2] (see Table 1).

---

[2] Following Murdock's (1958) rationale, *t* tests were used to evaluate statistical significance for percentage data.

Thus the present study replicated Twedt and Underwood's (1959) finding of a significant negative transfer effect with a C-B paradigm in a mixed list. Moreover, the magnitude of negative transfer was greatest in the A-Br paradigm, followed by the A-C and C-B paradigms, again replicating the findings obtained by Twedt and Underwood.

One of the main problems of this study concerned the covariation between amount of R-S learning and the amount of transfer in the C-B paradigm. It was decided that the most appropriate R-S measure for this purpose is the R-S recall score for C-B pairs on List 2. Although the selection of this measure restricted the range of individual differences (scores ranging from 0 to 3), it was thought that this score would best reflect the conflict in R-S associations postulated as a causal factor for negative transfer in the paradigm. That is, low scores would conceivably indicate a large degree of competition between B-A and B-C associations; high scores would conceivably indicate a resolution of the competition with the B-C associations dominating.

The obtained *r*'s between these R-S measures and transfer were .21 ($p < .10$) for absolute transfer and .35 ($p < .01$) for relative transfer. In both cases the positive correlation indicated greater negative transfer with decreasing R-S score. This tendency is further illustrated by arbitrarily dichotomizing the *S*s into a high R-S group on List 2 (scores of 3 and 2, $n = 37$) and a low group (scores of 1 and 0, $n = 23$). For absolute transfer the means for the high and low groups were −1.35 and −3.91, respectively ($t = 1.71$, $p < .10$); for relative transfer the corresponding means were −4.81% and −22.69% ($t = 2.72$, $p < .01$).

TABLE 1

Summary Statistics for List 2 S-R Transfer Effects

Paradigm	Absolute Scores (E − C)		*t* (from zero transfer)	Relative Scores (E − C/E + C)		*t* (from zero transfer)
	Mean	*SD*		Mean	*SD*	
C-B	−2.33	5.46	3.33	−11.67	26.08	3.46
A-C	−4.55	7.64	4.64	−16.30	27.90	4.53
A-Br	−7.40	7.34	7.87	−25.47	28.48	6.92

From the above results, it appears as if the presence of negative transfer in the C-B paradigm, relative to C-D, is a consequence of R-S learning, as implied by Twedt and Underwood. More generally, the prediction of transfer effects within any paradigm needs to take into account the relationships between all associations, both S-R and R-S, within the two lists. Whenever response competition, either in the forward or backward direction, is present between "responses" of the two lists, negative transfer is to be expected. In the A-C paradigm, competition occurs directly via S-R associations. In the C-B paradigm the competition is between S terms of List 1 with those of List 2 as responses to R terms common to both lists. As might be expected from the generally found smaller amount of R-S learning relative to S-R learning, the amount of negative transfer both in the Twedt and Underwood and present studies is greater in the A-C than in the C-B paradigm. As implied by Porter and Duncan (1953), the amount of negative transfer is especially great in the A-B_r paradigm where both S-R and R-S competition occurs. The present study replicated previous ones by finding the greater amount of negative transfer with the latter paradigm.

It is conceivable that the transfer effects noted above in the crucial C-B paradigm were influenced by the procedure involved in measuring R-S learning after List 1. In an attempt to evaluate the feasibility of this possible confounding effect, Group

TABLE 2
SUMMARY STATISTICS FOR R-S RECALL SCORES ON LISTS 1 AND 2

Paradigm	List 1		List 2		Difference Score (List 2 − List 1)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
C-D	1.80	.94	2.25	.84	.45	1.13
C-B	1.77	.89	1.83	.99	.06	1.13
A-C	2.03	.86	2.03	.88	0	.93
A-Br	2.00	.90	1.43	.98	−.57	1.10

C was compared with Group II of the experimental *S*s. These two groups were alike with respect to the content of the A-B and C-B pairs but differed with respect to R-S measurement after List 1. The mean relative transfer was -15.27% and -15.87% for Groups II and C, respectively. A t test did not permit rejection of the null hypothesis ($t = .06$). Thus, the presence of a List R-S measurement appeared to have little effect on negative transfer in the C-B paradigm.

R-S Transfer Effects[3]

Summary statistics for R-S recall scores are presented in Table 2 as determined for both lists and all four paradigms. An analysis of variance indicated no significant difference between the paradigms on List 1—between paradigm $MS = 1.11$, error $MS = .60$; $F(3, 177) = 1.85$, $p > .10$. However, an analysis of variance for List 2 yielded a significant paradigm effect—between paradigm $MS = 7.24$, error $MS = .63$; $F(3, 177) = 11.49$, $p < .001$. Separate t tests (with the error MS as the estimate of population variance) revealed significantly less List 2 recall for both C-B and A-Br pairs as compared to C-D pairs (t's of 3.00 and 5.86) and significantly less A-Br recall than either C-B or A-C (t's of 2.86 and 4.29).

In addition, an absolute transfer score was obtained for each S in each paradigm by subtracting his List 1 R-S score from his List 2 score. Summary statistics for the four paradigms on this score are given in Table 2. From Table 2 it is apparent that R-S recall increased for the C-D paradigm on List 2. A t test indicated that this increase departed significantly from the null value ($t = 2.87$, $p < .01$). For the A-Br paradigm, however, List 2 recall was significantly less than List 1 ($t = 4.07$, $p < .01$). For the C-B and A-C paradigms the difference in recall between the two lists was obviously not significant.

An analysis of R-S scores on List 2 for Group C revealed a paradigm effect generally comparable to that found for the main body of experimental *S*s on List 2. The means were 2.13, 1.80, 1.27, and .80 for the C-D, C-B, A-C, and A-Br paradigms, respectively, with the overall effect being significant as determined by analysis of variance—between paradigm $MS = 5.18$, error $MS = .58$; $F(3, 42) = 8.93$, $p < .001$. Again C-D yielded the greatest and A-Br the least amount of List 2 R-S recall. However, the C-B and A-C paradigms were reversed in order relative to the experimental *S*s.

In general, the R-S data suggest that R-S learning is susceptible to a learning-how-to-learn positive transfer effect much like that for S-R learning. The positive transfer from List 1 to List 2 is most apparent in the A-B, C-D paradigm where neither S-R not R-S associations from the two lists interfere with one another. As the interference increases, the amount of positive transfer would be expected to decrease. Generalizing from the present results, the interference is greatest in the A-Br paradigm where competing associations are present during the learning of List 2 for both S-R and R-S components. In fact, the negative effect from these competing associations seems to be significantly greater than the learning-how-to-learn transfer. For the remaining paradigms, the positive and negative transfer effects appear to be approximately equal, with zero transfer as the resultant.

---

[3] Since R-S associations are measured by the presentation of the R term, only the R-S recall trial data are included in this report. The free recall of stimuli was included in this study for a secondary purpose, unrelated to transfer effects. Actually, free recall and R-S trial recall, as indices of stimulus learning, indicated considerable agreement. For List 1 the means were 7.40 and 7.60, with an r of .56 ($p < .01$) between the measures; for List 2 the means were 7.60 and 7.54, with an r of .37 ($p < .01$).

References

HARCUM, H. R. Verbal transfer of overlearned forward and backward associations. *Amer. J. Psychol.*, 1953, **66**, 622–625.

MURDOCK, B. B., JR. "Backward" learning in paired associates. *J. exp. Psychol.*, 1956, **51**, 213–215.

MURDOCK, B. B., JR. Transfer designs and formulas. *Psychol. Bull.*, 1957, **54**, 313–326.

MURDOCK, B. B., JR. "Backward" associations in transfer and learning. *J. exp. Psychol.*, 1958, **55**, 111–114.

PORTER, L. W., & DUNCAN, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46**, 61–64.

STEVENS, S. S. *Handbook of experimental psychology.* New York: Wiley, 1951.

TWEDT, H. M., & UNDERWOOD, B. J. Mixed vs. unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58**, 111–116.

(Received April 9, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 206–211

RESISTANCE TO EXTINCTION AS A FUNCTION OF TEMPORAL RELATIONS DURING SENSORY PRECONDITIONING[1]

DELOS D. WICKENS AND HENRY A. CROSS
Ohio State University *Oklahoma State University*

This experiment investigated the effect of interstimulus interval within the range of 0 to 600 msec. in sensory preconditioning. The basic design was: 10 preconditioning trials to light and tone; GSR conditioning to light with shock UCS for 10 trials; 4 tests to tone. 4 groups of 20 Ss were used in different tone-light interstimulus intervals during preconditioning. The interval differences were 0, 100, 400, and 600 msec., tone preceding light. The groups differed significantly in response strength during the tests to the tone, with order of response magnitude from lowest to highest being 600, 0, 100, and 400; an order similar to that found in interstimulus interval curves in ordinary conditioning and in elements of a compound CS.

In an experiment by Wickens, Gehman, and Sullivan (1959) it was found that the long element of a stimulus compound acquires response strength according to temporal onset differences between the two elements that are generally congruent with previously reported researches investigating conditioning to a single stimulus, (Moeller, 1954; White & Schlosberg, 1952). The experiment indicated that the longer element of the compound stimulus showed an increased amount of response strength as its precedence time prior to the shorter stimulus—which always preceded the UCS by 500 msec.—increased from 100 msec. to 420 msec., but showed a decrement for a precedence time of 540 msec.

In interpreting this result it was assumed that some kind of sensory conditioning between the two conditioned stimulus elements had occurred. It was impossible to determine from that experiment whether the sensory conditioning arose from the fact that the shorter stimulus had acquired a secondary reinforcing characteristic from the UCS or whether it was a direct consequence of pairing the two conditioned stimuli.

In the present experiment the sensory preconditioning paradigm (Brogden, 1939) was employed, the independent variable being the time by which onset of one stimulus preceded the other in the preconditioning period. Although others have shown that time relations are important in sensory preconditioning (Hoffeld, Thompson, & Brogden, 1958; Silver & Meyer, 1954) the particular time relations used in the present experiment were chosen to conform with the above mentioned findings by Wickens, Gehman, and Sullivan (1959). The time values employed in this experiment are, therefore, quite different from those used in the other researches.

METHOD

Subjects.—A total of 80 men and women from an introductory course in psychology at the Ohio State University were used in the final analysis. Six *S*s were discarded because of failure to adapt to the stimuli during the preconditioning stage of the experiment. One *S* was discarded because of an unusually high resistance level.

---

[1] This research was supported in part by a grant, G6203, from the National Science Foundation.

Apparatus.—The time relationships between the stimuli were controlled by an arrangement of wave-form generators (Type 162), pulse generators (Type 163), and a power supply (Type 160A) obtained from Tektronix, Incorporated. The error in this equipment is less than 1 msec. The shock was produced by a dc power supply. One stimulus was a 1000-cps tone at approximately 40 db. re .0002 dyne/cm^2, produced by a Jackson oscillator and fed into earphones. The other stimulus was the illuminating of a 2-w. light behind a ruby panel glass located at eye level and about 4 ft. directly in front of *S*. Shock electrodes were fastened to the right hand. The GSR response was measured by means of a Fels dermohmmeter. This instrument permits measurement of the *S*'s resistance value at any moment in time. The recording electrodes were attached to the palm and back of the left hand. An electrode jelly composed of bacto-agar and zinc sulphate was used between the hand and the zinc electrodes.

The *S* sat in a dental chair in a semi-soundproof room 8 × 12 ft. with a ceiling of 8 ft. It was dimly illuminated by a small General Electric night light which was totally eclipsed by its adjustable plastic covering. The light was almost directly over the *S*'s head. A black curtain which hung from the ceiling formed a semicircle around *S*'s chair restricting his view of the rest of the room.

The stimulus-producing and recording equipment operated by *E* was located in an adjoining room.

Procedure.—The *S* was brought into the experimental room, seated in the chair, and after the attachment of the recording electrodes (left hand) and shock electrodes (right hand) was given directions. He was told that he was participating in a conditioning experiment and that in order to familiarize him with the stimuli he would first be given three mild electric shocks followed by several paired presentations of a tone and a light. After this the conditioning proper would start. During this time one of the previous stimuli (tone or light) would be paired with a mild electric shock to the right hand. The *S* was not told that on some later trials a stimulus would not be followed by shock.

After giving the instructions, *E* placed the earphones on *S* and left the room but did not close the door. The *S* was then warned to expect a shock. After the shock he was questioned as to its intensity and told that he would receive another one of slightly greater intensity. After this shock he was told that the next shock would be like the one actually employed in the experiment, namely 1.0 ma. The door to *S*'s room was then closed and the experimental procedure begun.

The procedure was the same for all groups except for time intervals between the two stimuli in the sensory preconditioning period, and was as follows:

1. Sensory preconditioning. Ten paired presentations of the light and tone were delivered. In all groups the light lasted 600 msec.; the duration of the tone prior to the onset of the light was the independent variable of the experiment. Four groups of 20 *S*s each, which differed with respect to the time interval variable, were run. In Group 1 the tone and light came on simultaneously. In Groups 2, 3, and 4 the precedence of the tone over the light was, respectively, 100, 400, and 600 msec. The tone and light terminated together so that the duration of the tone for the four groups was 600, 700, 1000, and 1200 msec. The intertrial interval varied in a prearranged random order with values of 30, 45, and 60 sec., having a mean interval of 42 sec.

2. Conditioning. The light, the second stimulus of the preconditioning session, was the conditioned stimulus for all groups. It lasted for 600 msec., and was followed by the shock 500 msec. after its onset. The two stimuli terminated together. The shock level was increased slightly and by nearly equal steps during the first 7 trials. The value on the first trial was 1.0 ma. and it reached 2.0 ma. on Trial 7. Intertrial intervals varied as in the sensory preconditioning period although a different random pattern was employed. A total of 10 reinforced trials was given.

3. Sensory preconditioning test. In the test stage of the experiment the tone alone was presented four times, and the intertrial intervals were the same as those employed during the first four preconditioning trials.

Response measure.—The resistance level was obtained at the time the CS was begun, and another reading was taken of the minimum value of the resistance immediately following the stimulation. Each of these values, obtained in ohms, was converted to a reciprocal and the difference between the two was obtained. This value was in turn multiplied by 10^7 and converted to a log score. The resistance values were read from a meter by *E* to units of 200 ohms.

RESULTS

Figure 1 shows the extinction curves of the four groups. It will be seen

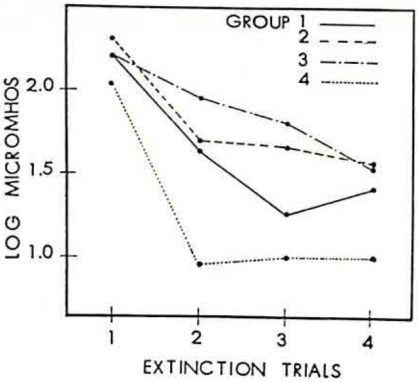


FIG. 1. Mean response magnitude for all groups during the four extinction trials.

that on Extinction Trial 1 the groups are quite similar in their response magnitudes. The *S*s were not warned of a test sequence to occur after the conditioning, and the clustering of all groups might be attributed to an alerting or perceptual disparity effect (Grings, 1960). By Trial 2, however, the groups arranged themselves as predicted and this was even more evident in Trial 3. By Trial 4 the differences lessened, presumably as a consequence of extinction.

An analysis of variance test suitable for repeated measures on the same *S*s (Edwards, 1956) was done and results are summarized in Table 1.

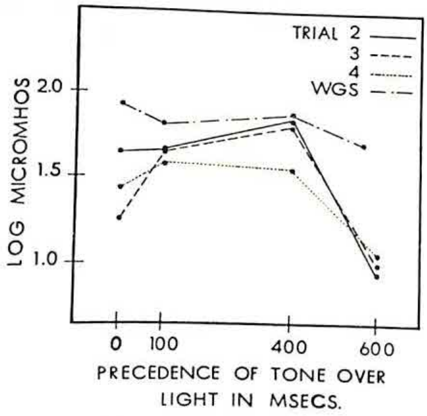


FIG. 2. Mean response magnitude as a function of interstimulus interval during the sensory preconditioning period.

It will be seen that the *F* (between trials) was significant ($p < .01$), indicating that extinction did indeed take place. Further, the groups gave responses of different magnitudes as indicated by the significant ($p < .05$) *F* for between groups. The *F* for interaction (Trials × Groups) was not significant.

Under the assumption that the Trial 1 data could be disregarded for the reason mentioned above, a single classification analysis of variance was done for Trials 2, 3, and 4 individually. Trials 2 and 3 both yielded significant ($p < .05$) *F*s, but Trial 4 proved to be nonsignificant. For Trials 2 and 3 *t* tests were computed. The mean scores for the various groups are represented in Fig. 2 which also includes the data for the combined Trials 2 and 3 of the Wickens, Gehman, and Sullivan (1959) study.

TABLE 1

ANALYSIS OF VARIANCE FOR EXTINCTION DATA (TRIALS 1–4)

Source	*SS*	*df*	*MS*	*F*
Between groups	16.2649	3	5.4216	2.86*
Between *S*s in groups	143.8439	76	1.8927	
Total between groups	160.1088	79		
Between trials	32.8637	3	10.8637	32.14**
Trials × Groups	5.2515	9	.5835	1.71
Pooled *S*s × Trials	77.7238	228	.3409	
Total within *S*s	115.8390	240		
Total	275.9478	319		

* $p < .05$.
** $p < .01$.

For Trial 2, two *t*'s were significant. Group 1 was significantly greater ($p < .05$) in response magnitude than was Group 4, as was Group 3 ($p < .01$). Group 2 also came near ($p < .10$) to establishing its superiority over Group 4. For Trial 3, Group 1 was inferior to 3 ($p < .05$); while Groups 2 and 3 were superior to 4 at the .05 and .01 levels, respectively. In summary, Group 4 is rather consistently inferior to any of the other groups while Group 3 tends to excell the other groups. The fact that a significant difference among the groups was not found on Trial 4 is not particularly surprising since the process of extinction would be expected to suppress and eventually equalize the performance level of all the groups.

DISCUSSION

It has been suggested that, since the experiment was begun with three shocks, some form of pseudoconditioning might be responsible for the differential in response magnitude among the groups in the postconditioning period. Therefore, two analysis of variance tests were conducted on the preconditioning data, a Trials × Groups analysis over the first four trials and also of the mean of the responses to Trials 9 and 10. The first analysis was chosen because it parallels the analysis used in the postconditioning, and the second because it is a measure of response strength to the conditioned stimuli just before conditioning. For the first analysis for trials $F(2, 152) = 9.74$ which is significant at the .001 level. Inspection of the data indicates that this F was due to a reduction in response magnitude with successive trials. However, the F for the interaction between groups and trials was .48 and that for groups was .79 and thus there is no evidence that the groups responded differentially at this time. The analysis for the mean of Trials 9 and 10 produced an F of .74, and again offers no evidence of difference in responsiveness among the groups at this time.

There is, therefore, no evidence in the data that either the initial shock testing nor the differences in temporal intervals employed during the preconditioning directly produced the differences in GSR magnitude obtained following the conditioning period. The alternate interpretation is that for the most part the response to the tone alone in postconditioning was a consequence of the interpolated conditioning period to the light plus the effects of the sensory preconditioning.[2] The disparity in response magnitude among the groups is assumed to be a reflection of a differential in effectiveness in sensory conditioning resulting from the interstimulus interval variable.

The experiment was undertaken primarily to determine whether the temporal parameter in sensory preconditioning would effect performance in a manner similar to its effect upon the elements of a compound stimulus. The results for the means score on Trials 2 and 3 in the Wickens, Gehman, and Sullivan study are presented in Fig. 2. In that experiment the group with a zero time difference was disregarded, and a curve was fitted to the data. This curve was complex, but it indicated that conditioning should be highly effective at the 400-msec. interval, poor at the 600-msec., and intermediate for the 100-msec. interval. It was on the basis of these findings that the time intervals for the various groups of the present experiment were chosen. The results of the present experiment conform closely to this prediction; the statistical analysis especially emphasizes the superiority of the 400-msec. group and the strikingly inferior performance of the 600-msec. group.

In view of the similar effect of time in these two experiments, one of which employed a sensory preconditioning design and the other an experiment in which the two conditioned stimuli were present at the time the UCS (reinforcement) occurred, it would seem justifiable to conclude that the occurrence of a reinforcement is not a necessary condition for establishing the effectiveness of the individual components of a compound stimulus.

---

[2] Since no control group was employed in this experiment, it is impossible to determine how much of the response magnitude is attributable to the sensory conditioning process itself rather than to other factors. The failure to include a control group was based upon the assumption that the phenomenon is well established and the fact that the experiment is concerned with relative differences in magnitude of sensory preconditioning effects rather than absolute amount.

This statement emphasizes the importance of the relationship of the two conditioned stimuli to each other rather than their individual relationship to the reinforcement. Actually there is some evidence at the molecular level that such relationships can be established. Thus Morrell and Jasper (1956) have shown that the cortical response to a visual stimulus may be elicited by a tone with which the light has been paired. The sensory preconditioning experiment simply adds a vehicle by means of which this association can be identified at a molar level.

Two other experiments have dealt specifically with time relations in sensory preconditioning, (Hoffeld et al., 1958; Silver & Meyer, 1954) and each of them has found the temporal variable to be effective. Silver and Meyer demonstrated that a forward order was superior to a simultaneous or backward order. Hoffeld, Thompson, and Brogden used five different values of a forward order, ranging from 0 to 4000 msec. Statistical analysis of their data indicated that their results could be described by an equation containing a quadratic as well as a linear term. This equation predicts a decrease in response magnitude from simultaneity out to about 560 msec. with progressive increases in response strength for the longer time intervals. If the 0- and the 560-msec. points were connected for the data of the present experiment a downward slope similar to the early portion of the Hoffeld, Thompson, and Brogden curve would be obtained. Therefore, despite differences in the type of *S* and technique, there seems to be an essential agreement between these two points in these two experiments. Since Hoffeld, Thompson, and Brogden included no groups between 0 and 500 msec., their experiment gives no information about the form of the curve within this interval. The data of the present experiment and the previous one using a compound stimulus indicates that the curve between these intervals is not monotonic. On the contrary, a quadratic term would be required to describe the trend of the data within the short span of 600 msec. In general this curve shows a form which is generally similar to that found in the studies of interstimulus intervals for ordinary conditioning (Kimble, 1947; Moeller, 1954; Reynolds, 1945; White & Schlosberg, 1952).

There is one contrast between the results of this study and that of the Wickens, Gehman, and Sullivan (1959) study which should be discussed. This deals with the performance of Group 1. In the present study it is slightly inferior to the 100- and 400-msec. groups whereas it was superior to all the groups in the previous study. Actually this result is to be expected, for in the study on the stimulus compound, the zero interval group is one for which both stimuli come on simultaneous 500 msec. before the UCS. It would be expected, therefore, that they could operate independently of each other and acquire their individual response strengths as a function of their individual relationship to the UCS. In the sensory preconditioning design, however, one stimulus is never paired with the UCS, so its response strength must be mediated by its association with the other stimulus. It would be expected that the zero time interval would be relatively ineffective, and such proved to be the case.

REFERENCES

BROGDEN, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1956.

GRINGS, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67**, 243–252.

HOFFELD, D. R., THOMPSON, R. F., & BROGDEN, W. J. Effect of stimuli time relations during pre-conditioning training upon the magnitude of sensory preconditioning. *J. exp. Psychol.*, 1958, **56**, 437–442.

KIMBLE, G. A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–115.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

MORRELL, F., & JASPER, H. H. Electrographic studies of the formation of temporary connections in the brain. *EEG clin. Neurophysiol.*, 1956, **8**, 201–215.

REYNOLDS, B. The acquisition of trace conditioned responses as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.

SILVER, C. A., & MEYER, D. R. Temporal factors in sensory pre-conditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 57–59.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

WICKENS, D. D., GEHMAN, R. S., & SULLIVAN, S. N. The effect of differential onset time on the conditioned response strength to elements of a stimulus complex. *J. exp. Psychol.*, 1959, **58**, 85–93.

(Received April 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 212–213

SUPPLEMENTARY REPORTS

A REPLICATION OF OPERANT CONDITIONING OF THE GSR[1]

ELLEN KIMMEL AND H. D. KIMMEL

University of Florida

An earlier study of operant conditioning of the GSR was replicated. 2 groups of 15 Ss were run in a totally dark room. After a 10-min. rest, the Contingent (C) group received a dim light following unelicited GSRs for 16 min., then 10 min. of extinction. The Noncontingent (N) group received the same number of lights per min. during reinforcement, when no GSRs were occurring. The dependent variable was the percentage of responses per min., using the average of the last 5 min. of rest as a baseline. Group C emitted significantly more responses during reinforcement and extinction than Group N ($p < .01$). The groups converged during extinction ($p < .05$). It was concluded that the unelicited GSR can be conditioned instrumentally.

Fowler and Kimmel (1962) have recently shown that the unelicited GSR can be conditioned instrumentally. Their findings are of sufficient theoretical significance that replication of their basic conditions in a new sample of *S*s is all but mandatory. The present study was done for this purpose.

Method.—The procedures of the two studies were identical except as follows: (*a*) The rest period prior to reinforcement was increased from 2 to 10 min., the last 5 of which were used as a basis for computing measures of percent response frequency. (*b*) The extinction period was lengthened from 4 to 10 min. (*c*) Two groups, instead of four, were run. Each contained 15 *S*s, matched as before with respect to the number of reinforcements per min. and, in addition, matched for sex. The two groups corresponded to the 16-min. Contingent and Noncontingent groups of the earlier study. (*d*) Due to the reduction in GSR amplitude occasioned by the use of a longer rest period initially, the criterion for reinforcing and counting unelicited GSRs was modified to include all responses detectable by the naked eye. Since these judgments were made "on the fly" during the reinforcement period for the Contingent group, a similar procedure was used for the Noncontingent group during reinforcement, *E* using a switch for counting purposes. (*e*) Responses occurring within 3 sec. following the reinforcement were not counted. In the earlier study, a 5-sec. "grace" period was used. (*f*) Ten *S*s were rejected due to *E* or equipment malfunction or for admissions of failure to follow instructions (moving, closing the eyes, etc.). These *S*s were approximately equally distributed between the two conditions (6 and 4). (*g*) The duration of the visual reinforcing stimulus was reduced to 0.1 sec. but its other physical properties were approximately the same as those described in the earlier report (i.e., a $\frac{1}{4}$-in. square dim white light about 30 in. from *S*'s face) (Fowler & Kimmel, 1962).

Results.—Figure 1 shows the average percentages of transformed response frequency during reinforcement and extinction, in 2-min. blocks. The figure indicates that the two groups separated at the first 2-min. point and remained apart thereafter. The Contingent group responded above the 100% level throughout. The overall difference between the two conditions during reinforcement was highly significant, $F(1, 28) = 8.99$. No interaction between Contingency and Minutes was observed. Analysis of the extinction data showed that both the Contingency and the Contingency × Minutes effects were significant, $F(1, 28) = 5.51$ and $F(4, 112) = 2.69$.

Discussion.—The results confirm and strengthen the conclusion that the GSR can be conditioned instrumentally. Lengthening the initial rest period had the desired effect of making the difference between conditions more pronounced than in the earlier study and, in addition, raising the Contingent group above 100%. Increasing the extinction period made it possible for the interaction between Contingency and Minutes to reach significance, as well as emphasizing the upward trend in the Noncontingent group.

---

[1] Based upon a Master of Arts thesis at the University of Florida. Partial support came from Grant NIH-M-6060.

Taken together, the two studies provide strong evidence that autonomically mediated responses can be conditioned instrumentally.

The postexperimental interview was used to ensure that only data from *S*s who were unable to verbalize "awareness" of the contingency of reinforcement upon "something they might have done" would be employed. While it is true that some *S*s may have been making a somatically mediated response about which they could not verbalize, the likelihood of this possibility is attenuated somewhat by the results of the Noncontingent group. Here the *S*s would have had to learn to *not* make some somatically mediated response during training and, then, during extinction, to extinguish this tendency to *not* respond. It would undoubtedly be of value to obtain simultaneous records of several somatically mediated responses to determine whether the occurrence of any were correlated with the occurrence of the unelicited GSR. We are able, at present, to rule out only movement of the hand holding the GSR electrodes, since the observed responses are in no way qualitatively similar to such "motor artifacts." It would appear that the burden of (negative) proof is on those who suspect some somatically mediated reaction or other, since our conclusion that autonomically mediated responses can be conditioned instrumentally can be challenged on the basis of an almost infinite number of possible somatic mediators.

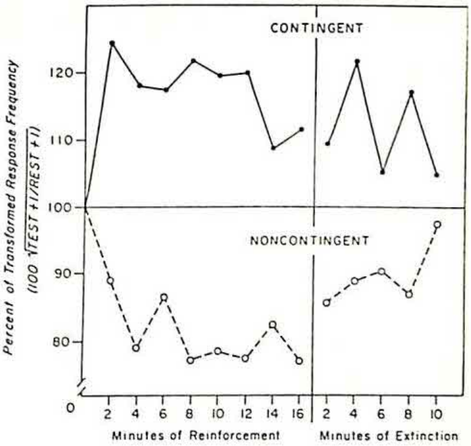


FIG. 1. Percentage of transformed response frequencies during 16 min. of reinforcement and 10 min. of extinction, in 2-min. blocks.

REFERENCE

FOWLER, R. L., & KIMMEL, H. D. Operant conditioning of the GSR. *J. exp. Psychol.*, 1962, **63**, 563–567.

(Early publication received August 25, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 213–214

EFFECT OF RUN LENGTH ON THE "GAMBLER'S FALLACY"[1]

PETER L. DERKS

College of William and Mary

Event schedules with short, long, and extra long event runs but with 75:25 event proportions were predicted by 3 groups of 20 Ss. Short runs resulted in negative recency. Both long run groups showed positive recency after 100 trials.

Several investigators have found that the successive occurrence of an event increases then decreases the probability that *S*s will predict that event. This phenomenon has been called the "gambler's fallacy" or negative recency.

There is intuitive justification for this gambler's fallacy. For example, in a finite n element series with a subpopulation of A with $A = 3/4\ n$ (75:25 schedule) the median run length of A is 3. When A occurs once or twice in succession it will occur again more than half the time while less than half the A runs are greater than 3. The proportion of A predictions on a 75:25 schedule has been found to reach a maximum after three successive A occurrences and then decreases.

However, other experimenters report that

---

[1] This research was supported in part by Grant A-158 from the College of William and Mary and Public Health Service Research Grant M-5193 (A) from the National Institute of Mental Health. The author wishes to thank Louis Gurnee Blanchard and Mary Ingles for their aid in experimentation and data tabulation.

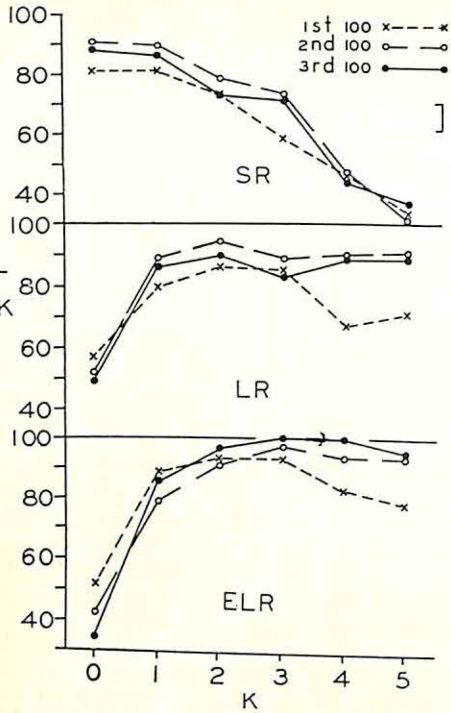


FIG. 1. T_k as a function of k, vertical bar = 2 estimated $SD \cong 9\ T_k$. (See Derks, 1962, for estimation of *SD*.)

the probability of prediction of an event continues to increase with its occurrence. One reason for this disparity is the difference in the length of the experiments. As *S*s continue to predict the schedule negative recency becomes positive recency (Derks, 1962; Edwards, 1961; Lindman & Edwards, 1961).

Event run length also seems to be an important factor. In this experiment *S*s predicted 75:25 schedules of different run lengths. The progressive effect of these different schedules on the gambler's fallacy was examined.

Method.—The binary events were nonsense syllables ZEJ and ZOJ (Davidson, Suppes, & Siegel, 1957) typed on 3 × 5 in. cards. There were 300 cards in the deck with 225 ZEJ. Short run (SR), long run (LR), and extra long run (ELR)[2] event schedules were constructed in 75:25 proportion. The run length distribution was maintained for each of three 100-trial blocks, but the run length order was random. Schedule SR had no A runs of greater than 5. Schedule LR had A runs from 2 to 11 with no single A events. Schedule ELR had no A runs of less than 5 with median run lengths 6 and 7.

Sixty undergraduates at the College of William and Mary served as *S*s. There were three groups of 20 *S*s with each *S* tested on only one schedule. The *S* was instructed to predict whether each successive card was ZEJ or ZOJ, and to try to be correct as often as possible. The *E* recorded the responses and *S* looked at the syllable on the card. The experiment ended when *S* had predicted and seen all 300 cards.

Results.—The prediction of A runs was examined by T_k as described in an earlier study (Derks, 1962). A sample of five occasions on which A occurred five times in succession was selected for each 100-trial block of each schedule. The A predictions were summated for each successive occurrence of A (k) for each group. Figure 1 gives T_k as a function of k for the three successive 100-trial blocks of the three schedules.

Schedule SR resulted in negative recency. The LR and ELR schedules produced positive recency after the first 100 trials. Thus event run length is an important determiner of the fine grain of human binary prediction.

The progressive alteration of the gambler's fallacy occurs relatively early for long runs. The increase in prediction of A after it has appeared four and five times in succession occurs prior to Trial 200 for the LR and ELR schedules. Changes in prediction of short runs of events apparently take much longer.

REFERENCES

DAVIDSON, D., SUPPES, P., & SIEGEL, S. *Decision-making: An experimental approach.* Stanford: Stanford Univer. Press, 1957.

DERKS, P. L. The generality of the "conditioning axiom" in human binary prediction. *J. exp. Psychol.*, 1962, **63**, 538–545.

EDWARDS, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

LINDMAN, H., & EDWARDS, W. Supplementary report: Unlearning the gambler's fallacy. *J. exp. Psychol.*, 1961, **62**, 630.

(Received March 3, 1962)

---

[2] Group ELR was suggested by Stanley Williams.

Journal of Experimental Psychology
1963, Vol. 65, No. 2, 215–216

TWO METHODS FOR TESTING SERIAL MEMORIZATION[1]

NANCY C. WAUGH

Harvard University

The learning curve obtained with the method of serial recall (alternate viewing and recall of an entire list) is compared with that obtained with the classical method of anticipation. Both are rectilinear with almost identical slopes. Each can be interpreted to represent a cumulation of memory spans over trials, in accordance with the results of previous studies.

The curve for serial learning has been found to be rectilinear (Waugh, 1961) when *S* alternately views an entire series and attempts to reproduce it.[2] Unlike the traditional method of anticipation, this procedure has the practical advantage of being suited to group experiments. It is possible, however, that the two methods elicit from *S* different strategies and hence different rates of learning.

With the method of serial recall, *S* is not required to associate each item with its predecessor on every trial. Instead, he can rehearse a fixed number of new items and ignore the rest each time he goes through a list. The linear learning curve would then represent this fixed number of items multiplied by the number of trials. When recall is tested by the method of anticipation, on the other hand, we assume that *S* pays equal attention to every item on every trial, since attempting to recall one item on cue would presumably preclude rehearsing another. Must *S* therefore partially learn many items, rather than completely memorizing a few, on each trial? Will his learning curve accordingly fail to be linear? The purpose of this study was to determine whether the two methods, serial recall and anticipation, do result in different kinds of learning functions.

Method.—Four *S*s, paid volunteers, each learned 20 lists of 60 monosyllabic English words, to give a total of 80 *S* lists. The construction of these lists has been described elsewhere (Waugh, 1961). Ten of them were reproduced by the method of anticipation, while the other 10 were reproduced by the method of serial recall. Each was exposed and recalled for six trials. Every *S* saw the lists in a different random order, over the course of five individual experimental sessions.

The words, which were typewritten on index cards, were shown one at a time through a window in a screen that separated *E* and *S*. When learning was by the method of anticipation (Cond. A), each word was shown for 2 sec. and was followed by a blank interval of 2 sec. The *S* pronounced the word out loud when it was exposed and then attempted to anticipate the next word during the blank interval. Under the method of serial recall (Cond. R), the words were exposed in immediate succession, each for 2 sec. The *S* was again required to read each word aloud. At the end of a series, he was given 2 min. in which to recall the series orally. The total time allowed for the exposure and recall of a list was therefore 4 min. per trial under each condition (4 sec. for each of 60 words). There was a 4-sec. interval between trials.

Results.—The *S*'s responses on each trial under Cond. A were scored in terms of the number of words correctly anticipated. Under Cond. R, the responses were scored in terms of the number of successive *pairs* of words correctly recalled on each trial, regardless of absolute serial position. A word was therefore considered to have been recalled correctly if the response which came before it was the word which actually preceded it in the list. For example, if the sequence ABCDE was presented and *S*'s response was DEABC, he was said to have recalled three pairs: DE, AB, and BC.

The group learning curves based on these scores are shown in Fig. 1. The upper set of points represents the mean number of words correctly anticipated under Cond. A, estimated across *S*s and lists ($N = 40$).[3] The lower set represents the corresponding mean number of pairs recalled correctly under Cond. R, trial by trial. The lines fitted to these points by a method of least squares are $\mathrm{N} = .6 + 8.3\ \mathrm{T}$ for Cond. A and $\mathrm{N} = -3.8 + 8.2\ \mathrm{T}$ for Cond. R, where N is the number

---

[1] This investigation was supported by a grant from the Carnegie Corporation of New York for the Harvard Center for Cognitive Studies. David Otto and Peter Stephenson gathered the data and helped analyze them.

[2] This procedure, which has been described as a modified method of complete presentation (Waugh, 1961), will be referred to here as the method of serial recall.

[3] The value shown for $\mathrm{T} = 6$ is a median rather than a mean, since the distribution of scores on this trial was noticeably skewed.

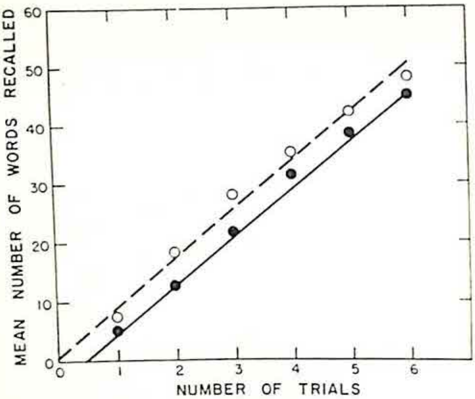


Fig. 1. Learning curves obtained under the methods of anticipation (open circles) and serial recall (closed circles).

of words recalled and T is the number of trials.

The deviations of each set of data from its fitted function were tested for significance by an analysis of variance of the mean recall scores obtained by each *S* on each trial. When the *MS* for deviations from linear regression was tested against the within-trials *MS*, the resulting *F* was less than unity, for both Cond. A and R. The variance of the scores, therefore, is in both cases largely accounted for by their linear regression on trials. We accordingly have no grounds for rejecting the hypothesis that both learning curves are rectilinear.

Note that the null hypothesis that was tested states simply that the learning curves are rectilinear, not that they are identical. There was no reason to assume beforehand that they would have the same slope or intercept. We might in fact suppose that recall would be easier under Cond. A than under Cond. R, since *S* is provided with overt cues in the former case but not in the latter. This appears to be true only on Trial 1. The number of items recalled on that trial is about 8.5 under Cond. A and about 4.5 under Cond. R. (This disparity may actually reflect the different scoring procedures followed with the two sets of data.) On each subsequent trial, there is an increment of about eight pairs under either condition. If we assume that a pair is seldom forgotten once it has been recalled, then this constant increment of 8 represents the number of new pairs learned per trial. The data shown in Fig. 1 are thus consistent with the hypothesis (Waugh, 1962) that only a memory span's worth of items can be learned within a given interval of time, and that the linear learning curve simply represents the summation of spans over trials.

The hypothesis was put forth that the method of serial recall perhaps allows *S* to attend to just a few items on each trial, while the method of anticipation may require him to rehearse all of the items in the list. Since we find that lists are learned at about the same rate under either method, the hypothesis may be incorrect: *S* may in fact rehearse selectively under both Cond. A and R, or he may rehearse uniformly in both cases. Or he may rehearse selectively under the one method and uniformly under the other, without its making any difference. In order to choose among these hypotheses, we shall have to discover either how to regulate *S*'s attention, or how to determine what he is rehearsing.

References

Waugh, N. C. Free versus serial recall. *J. exp. Psychol.*, 1961, **62**, 496–502.

Waugh, N. C. Length of series and the learning curve. *Amer. J. Psychol.*, 1962, **75**, 177–192.

(Received March 23, 1962)

Journal of
Experimental Psychology

Vol. 65, No. 3 March 1963

EFFECTS OF TASK COMPLEXITY AND TASK ORGANIZATION ON THE RELATIVE EFFICIENCY OF PART AND WHOLE TRAINING METHODS [1]

JAMES C. NAYLOR AND GEORGE E. BRIGGS

Ohio State University

Task complexity (2 levels), task organization (independence of task dimensions—2 levels), and training method (whole vs. progressive-part) were combined factorially in an 8-group transfer of training study of skill in a Markov prediction task. A hypothesized interaction of the 3 independent variables was supported by the data, thus suggesting the following training principle: for a relatively highly organized (integrated) task, a whole-task training method should be superior to a part schedule at all levels of task complexity; however, for a relatively unorganized task (all task dimensions independent), an increase in task complexity will result in a part-task training schedule becoming superior to whole training.

In an attempt to resolve much of the confusion in previous studies on part vs. whole training methods, Naylor (1962) has suggested a task taxonomy by which to systematically categorize several task parameters as determinants of method effectiveness. The basic structure of the taxonomy states that task difficulty, as defined by the relative performance levels of *S*, is a function of both task complexity and task organization. Task complexity is defined as the demands placed on *S*'s information-processing and/or memory-storage capacities by each of the task dimensions independently, while task organization refers to the demands imposed on *S* due to the nature of the interrelationship existing among the several task dimensions.

It is interesting to note that whereas an increase in task complexity usually results in an increase in task difficulty, an increase in task organization can either increase or decrease task difficulty, depending upon the nature of the interrelationships used to define

---

[1] This research was carried out in the Laboratory of Aviation Psychology and was supported by the United States Navy under Contract No. N61339-950, sponsored by the United States Naval Training Device Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part for any purpose of the United States Government.

task organization. For example, Briggs and Waters (1958) manipulated task organization by varying the degree of interaction present in a two-dimensional tracking task, and they found increased task difficulty as task organization (the amount of interaction) was increased; however, in most verbal learning tasks the presence of relationships between dimensions usually results in a decrease in task difficulty (Lyon, 1914; McGeoch, 1930).

In his article Naylor (1962) suggests that training method effectiveness should be related to both task organization and task complexity, plus he hypothesizes a possible interaction between the two variables. A recent evaluation of the effect of task complexity on method efficiency (Briggs & Naylor, 1962) obtained results which failed to either directly support or deny the relationship. Since Naylor's interaction hypothesis would appear to be the most critical relationship, Briggs and Naylor have suggested that this may have caused the ambiguity found in the study evaluating task complexity alone.

The present study was designed as a test of the Naylor interaction hypothesis: there should be an interaction between task complexity, task organization, and training methods such that (*a*) for tasks of relatively high organization as task complexity is increased, whole-task training should become relatively more efficient than the part-task methods; however, (*b*) for tasks of relatively low organization (all task dimensions being independent) an increase in task complexity should result in a part-task method becoming superior to the whole-task method. The two training methods compared were a progressive-part-task schedule and a whole-task practice method.

METHOD

There were two levels of task complexity, two levels of task organization, and two training methods (progressive-part and whole) incorporated in the design. Table 1 provides a summary of the conditions. There were 14 *S*s per group and none had experienced a similar laboratory task prior to service. There were five daily sessions of approximately 1-hr. duration each. Training occurred over the first 3 days and transfer was effected on the last two sessions in a transfer of training paradigm.

Subjects and tasks.—The *S*s were 112 undergraduate female students in introductory psychology classes. They were assigned at random to the eight groups subject only to the restriction that all groups be of equal size. All *S*s were paid $7.50 for the entire series of five 1-hr. sessions.

The task was basically a Markov prediction situation which required *S* to predict three aspects of an event just prior to the occurrence of the event. The event was a visual stimulus projected on a screen and a series of such stimulus events occurred at the rate of one event per 18 sec. The several discrete events varied in terms of three stimulus dimensions: (*a*) type of stimulus—aircraft, carrier, or submarine; (*b*) location of stimuli—left (L), center (C), or right (R); and (*c*) number of stimuli—one, two, or three objects of a single type. Thus, a given stimulus event consisted of one of three vehicle types in one of three possible locations or zones on the display and in number ranging from one to three vehicles.

Series of stimulus events were photographed on 16-mm. film. There were 60 stimulus events per series and a complete series represented one trial. There were two trials per day. Instructions to *S* emphasized

TABLE 1

EXPERIMENTAL CONDITIONS DEFINING THE EXPERIMENTAL GROUPS

Group	Training Method	Task Complexity	Task Organization
1	Whole	High	High
2	Progressive	High	High
3	Whole	High	Low
4	Progressive	High	Low
5	Whole	Low	High
6	Progressive	Low	High
7	Whole	Low	Low
8	Progressive	Low	Low

that she was to predict *what, where,* and *how many* stimuli would appear and that she was to mark down her predictions on special record sheets prior to the occurrence of each stimulus event. The *S* was instructed also to look for any relationships *within and between* the stimulus dimensions because any relationship discovered would help in the predictions since, although there were four different series representing each experimental condition, the basic relationships between and within stimulus conditions did not change from trial to trial for a given group of *S*s.

Since the criterion task was a Markov prediction situation, it was possible to manipulate both task organization and task complexity by varying the sequential dependencies of the inter- and intradimensional events, respectively, as follows:

Task complexity was defined as the degree of predictability present in each of the task components considered in isolation (intradimensional predictability); thus, this parameter was varied by modifying the predictability *within* each of the separate stimulus dimensions. The digram frequencies of a stimulus series were adjusted while keeping the monogram frequencies constant and equal at all times. Thus, for the zone dimension, all three zones were used an equal number of times in each series (20 stimulus events). However, for the highly complex series all nine digrams (LR, LC, LL, CL, CC, CR, RL, RC, RR) were also equated for frequency of occurrence, while for a low-complexity series the digrams LL, CC, and RR never occurred, i.e., the same zone was never used twice in succession. This reduced the average within-dimension information from 3.160 bits in a high-complexity series to 2.577 bits for a low-complexity series. Maximum possible information for independent stimulus events, considering only monograms and digrams, would be $\log_2 9$ or 3.169 bits. Considering *all* higher-order terms, maximum information possible per event is $\log_2 59$, or 5.88 bits.

As indicated above, the definition of task organization is the degree to which separate task dimensions have meaningful relationships with each other (interdimensional predictability). This parameter was varied by modifying the predictabilities existing between dimensions, e.g., the prediction of zone from information concerning type. For the stimulus series of low organization, the cross-dimensional digrams were all equal in frequency, creating essentially a complete independence of dimensions. There are, with three dimensions, a total of six potential cross-dimensional relationships which might be established. These are predicting (*a*) type from zone, (*b*) type from number, (*c*) zone from number, (*d*) zone from type, (*e*) number from type, and (*f*) number from zone. For lists of high organization, predictability was established across dimensions using two of the possible six relationships: number was made predictable on the basis of zone information and zone was predictable from type information. This was done in each case by eliminating six of the nine possible cross-dimensional digrams for these two relationships. This resulted in series having an average cross-dimensional information metric of 2.449 bits for the highly organized series as compared with 3.119 bits for the low-organization series. Table 2 summarizes the information metrics describing each of the four task conditions.

TABLE 2

INFORMATION PER STIMULUS (IN BITS) DEFINING EACH SERIES

	High Organization		Low Organization	
	High Complexity	Low Complexity	High Complexity	Low Complexity
Average intralist information	3.160	2.577	3.160	2.577
Average interlist information	2.449	2.449	3.119	3.119

These variations resulted in a total average information (summing across both the within and between dimensions) of 2.74 bits for high-organization, high-complexity; 2.50 bits for high-organization, low-complexity; 3.14 bits for low-organization, high-complexity; and 2.93 bits for low-organization, low-complexity conditions.

Procedure.—The groups having whole practice (Groups 1, 3, 5, and 7) made all three predictions (type, zone, and number) for every stimulus event throughout the entire 5-day period. The groups having progressive-part practice (Groups 2, 4, 6, and 8) predicted only type on Session 1, type *and* number on Session 2, number *and* zone on Session 3, and all three dimensions on Sessions 4 and 5. Thus, practice for these *S*s stressed those cross-dimensional relationships that led to maximum predictability, and Groups 2, 4, 6, and 8 can be considered as experimental

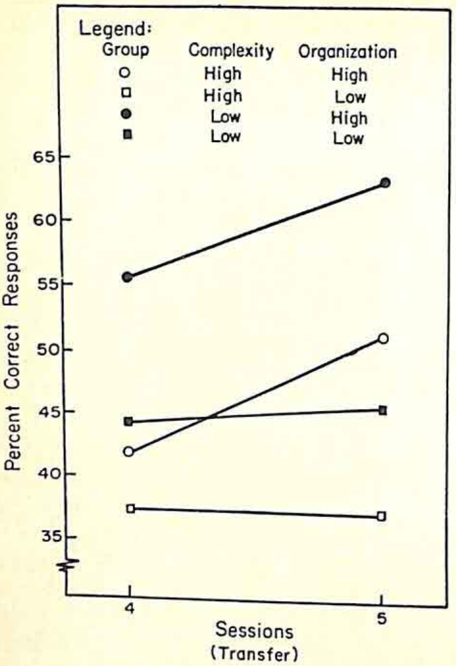


FIG. 1. Mean number of correct responses during Transfer Sessions 4 and 5 as a function of task complexity and task organization.

groups in a transfer of training paradigm while Groups 1, 3, 5, and 7 are control groups.

The total number of correct predictions per 60-stimulus-event trial was used as the dependent variable. In order to make the data comparable during the first 3 days when part training was in progress, *percent* correct predictions were calculated and employed in the data presentation.

RESULTS

It was necessary to ascertain the extent to which manipulation of task complexity in the manner described did indeed affect task difficulty. Figure 1 shows the average performance of *S*s on the two transfer sessions (Sessions 4 and 5) for each of the four task conditions. Each point is the average of both the whole and the progressive-part groups assigned to that particular task. It is clear from the figure that as task complexity was increased, task difficulty also was increased, as measured by total task performance. Thus, for the highly organized task, the more complex version resulted in a smaller number of correct responses. The same was true of the less highly organized task.

The use of the information metric as a useful means of quantifying both task organization and task complexity (and thus predicting task difficulty) appears to be very meaningful. This is illustrated by Fig. 2, which shows how the changes in average relative information of the four task conditions affect task performance. The high-complexity, low-organization series were used as a base. Thus, it appears that intradimension information provided a good index of complexity, average interdimension information provided a good measure of organization, and average intra- and interdimension information provided a good index of total task difficulty.

The results from the training and transfer sessions are summarized in Fig. 3 (high-complexity groups) and Fig. 4 (low-complexity groups). The first three sessions are training and the latter two are transfer conditions for Groups 2, 4, 6, and 8.

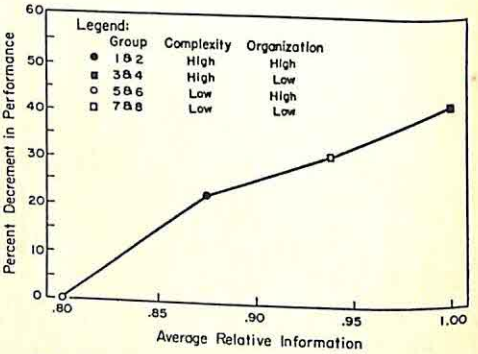


FIG. 2. Percent decrement in performance (number of correct predictions) as a function of the average relative information in a stimulus series. (The high-complexity, low-organization series was used as a base in computing average relative information.)

Training.—The use of sequential dependencies within a stimulus dimension and between stimulus dimensions (to define task complexity and task organization, respectively) appears to have separated groups as predicted: the worst performance at the end of training occurred under the high-complexity, low-organization condition (Groups 3 and 4) and the best performance was attained under the low-complexity, high-organization arrangement (Groups 5 and 6). Groups 3 and 4 could find no relationships between or within the stimulus dimensions of the event series because none were present; however, there were helpful relationships *both* between and within the stimuli for the series viewed by Groups 5 and 6, and their superior performance at prediction was the result. The remaining four groups could discern relationships either within or between stimuli in the series (but not both) and their performance levels, as expected, were intermediate to those of the extreme conditions.

Transfer.—There are a number of ways to express the transfer performance as shown by Sessions 4 and 5 of Fig. 3 and 4. One of the more meaningful ways is to express the transfer performance of an experimental group relative to its control group. Further, it is logical to make the comparison relative to the improvement of the two groups from initial training to initial transfer performance. An index to express this is $(C_i = E_T/C_i - C_T) \times 100$, where C_i is the performance of the appropriate (whole task) control group during initial training, C_T is the performance of the same group during initial transfer, and E_T is the performance of the experimental group during the same transfer trials. To the extent that the index is less than 100% for a given experimental group, to that extent the particular training method experienced by that group was less efficient than the whole-task training method. The results of applying this index are shown in Table 3.

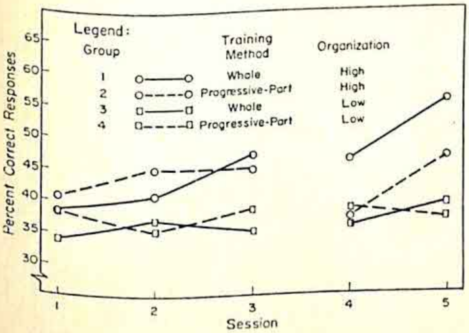


FIG. 3. Training and transfer performance for groups transferring to the high-complexity level task.

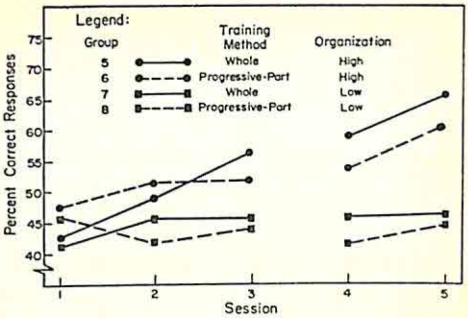


FIG. 4. Training and transfer performance for groups transferring to the low-complexity level task.

Note first the low-organization indices: higher relative transfer occurs for the more complex task (233.3%) than for the less complex task (27.2%).

TABLE 3

TRANSFER INDICES FOR GROUPS 2, 4, 6, AND 8

Task Organization	Task Complexity	
	Low	High
Low	27.2%	233.3%
High	67.7%	−20.0%

TABLE 4

Analysis of Variance of Transfer Sessions Data (Days 4 and 5) for All Eight Groups

Source	*df*	*MS*	*F*
Training methods (M)	1	12,705	5.717*
Organization (O)	1	81,277	36.578**
Complexity (C)	1	75,488	33.972**
M × O	1	5,333	2.400
M × C	1	102	.040
O × C	1	5,893	2.650
M × O × C	1	13,812	6.216*
Ss within groups (*Ss*/G)	104	2,222	
Transfer sessions (S)	1	16,646	52.677**
S × M	1	8	.025
S × O	1	9,845	31.145**
S × C	1	28	.089
S × M × O	1	24	.076
S × M × C	1	383	1.212
S × O × C	1	128	.405
S × M × O × C	1	594	1.880
S × *Ss*/G	104	316	

* $p < .05$.
** $p < .01$.

However, now note the striking reversal when one compares the two indices for the highly organized task: here, the higher level of relative transfer occurs for the less complex task (67.7%) and the more complex task resulted in less (even *negative*) transfer (−20%). Thus, whereas a progressive-part training method was much worse than a whole-task method under conditions of high task organization (and high complexity), the same method was considerably better than the whole-task method under conditions of low task organization (and high complexity). Naylor's (1962) prediction is borne out thereby in the data: task complexity and task organization *do* interact to influence performance *and* to determine what training method will be most appropriate for the acquisition of skill.

Table 4 is a summary of the results of an analysis of variance performed on the transfer data. Of interest is the fact that the differences due to training method, task organization, and task complexity were all statistically significant at $p < .05$. These results can be seen in Fig. 3 and 4: on the average, groups trained by the whole method (Groups 1, 3, 5, and 7) were superior to the progressive-part groups (Groups 2, 6, and 8) during transfer, with the most notable exception of the progressive-part group (Group 4) which experienced the high-complexity, low-organization condition; on the average, transfer performance was higher on the less complex (Groups 5–8) than on the more complex (Groups 1–4) task; and, in general, the high-task-organization condition (Groups 1, 2, 5, and 6) resulted in performance superior to that under the low-task-organization condition (Groups 3, 4, 7, and 8).

The significance of the Methods × Complexity × Organization interaction was the most notable finding of the analysis. This was indicated previously from the results of the transfer indices (Table 3) where an interaction is clearly apparent. This interaction was the essential test of the hypothesis under investigation, indicating that the choice of a training method is dependent upon *both* task complexity and task organization.

The significant variance in Table 4 due to sessions shows the rapidity of adjustment during transfer to the new task, and the Sessions × Organization interaction indicates that this rapidity differs as a function of the organizational characteristics of the task. Thus, the highly organized tasks are subject to rapid behavioral adjustment as opposed to only slight changes in performance with the low-organization tasks.

DISCUSSION

Several interesting points emerge in examination of the results. The relative transfer indices showed that the progressive-part training group actually exceeded the whole training group in transfer performance on the high-complexity, low-organization task. From Fig. 1 it is evident that this version of the task was by far the most difficult of the four task conditions used in the study. Thus, for an unorganized task, part practice surpassed whole practice in efficiency when the task was complex. While it is somewhat refreshing to find a condition under which whole training was not superior to another training method, the result was not unexpected. Prior research (Briggs & Naylor, 1962) has shown that increased complexity (and thereby difficulty) increased the size of the progressive-part transfer index from 75% to 91%, the latter group nearly equaling the whole-task group in transfer efficiency. The task used in the previous study was a three-dimensional tracking problem which involved no interaction between components and could therefore be considered to be of low organization. Therefore, the high-complexity, low-organization version of the task used in the present study must have exceeded sufficiently in difficulty the most complex version of the task used in the previous study so that progressive-part training did indeed result in greater transfer performance than the whole method.

The hypothesis that the relative efficiency of the two training methods would be related to an interaction between task complexity and task organization seems substantiated. Both the absolute differences in performance (as indicated by the analysis of variance of Table 4) and the relative differences in transfer (as indicated by the transfer scores of Table 3) show the same interactive pattern. Given a task of high organization, increasing the difficulty (via increasing complexity) tends to favor whole-practice efficiency relative to progressive-part efficiency. However, as was pointed out above, exactly the reverse occurs with tasks of low organization. Here, increasing difficulty results in increased relative effectiveness of the progressive-part method.

These results are logical. By definition, a task of low-level organization consists of essentially independent components or dimensions, and *S* need not be concerned with the learning of relationships or intercouplings *between* task components. Instead, skill acquisition consists primarily of learning and applying the rules of information processing and/or memory storage for each dimension separately. Therefore, an increase in task complexity results in increased component task difficulty, the individual components still being independent, and it is logical that a progressive-part training method would be efficient since this method emphasizes the acquisition of skill in dealing with the individual components of the task. However, for tasks of higher levels of organization, *S* cannot treat the several task dimensions as separate and individual subtask requirements since, by definition, they are interrelated, and it is important that he acquire skill not only on the specific information-processing/memory demands of the component tasks, but also that he develop time-sharing and other intercomponent skills as required by the cross-dimension relationships. The whole-task training method should be more efficient under these conditions since this method emphasizes the acquisition of skills in dealing with the total (integrated) task.

For a relatively highly organized task, the whole training method should be superior to a part-task schedule regardless of the level of task (individual component) complexity. This follows since manipulations of individual component complexity can only add to or at least leave unchanged the cross-component relationships, and thus *S* must deal with total task demands regardless of the level of component complexity.

Seymour (1956) has suggested that a part-task method should become superior to the whole-task training method for tasks requiring complex perceptual skills (the high-complexity conditions as defined in the present report). The present data indicate that this conclusion is valid only for tasks of relatively low organization.

REFERENCES

BRIGGS, G. E., & NAYLOR, J. C. The relative efficiency of several training methods as a function of transfer task complexity. *J. exp. Psychol.*, 1962, **64**, 505–512.

BRIGGS, G. E., & WATERS, L. K. Training and transfer as a function of component interaction. *J. exp. Psychol.*, 1958, **56**, 492–500.

LYON, D. O. The relation of length of material to time taken for learning. *J. educ. Psychol.*, 1914, **5**, 1–9.

MCGEOCH, J. A. The influence of associative value upon the difficulty of nonsense-syllable lists. *J. genet. Psychol.*, 1930, **37**, 421–426.

NAYLOR, J. C. Parameters affecting the relative efficiency of part and whole practice methods: A review of the literature. *USN Train. Dev. Cent. tech. Rep.*, 1962, No. 950-1.

SEYMOUR, W. D. Experiments on the acquisition of industrial skills. Part 3. *Occup. Psychol.*, 1956, **30**, 94–104.

(Received April 11, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 225–232

RESPONSE INTERFERENCE IN PAIRED-ASSOCIATE LEARNING[1]

LEONARD M. HOROWITZ AND SUZANNE R. LARSEN
Stanford University

Response familiarization can facilitate or hinder subsequent paired-associate (PA) learning, but the critical task-variables have not been known. The present study suggests that items learned during familiarization may compete with the responses of the PA task. 80 Ss participated in a familiarization task and then learned English-Japanese word pairs by the PA or associative matching method. (In associative matching S does not produce the responses, so extralist responses cannot intrude.) Group I was familiarized with only relevant responses; Group II, with irrelevant but similar responses; Group III, with irrelevant, dissimilar responses; and Group IV had no familiarization. In associative matching the performance of Groups I and II was superior to that of Groups III and IV ($p < .001$). In PA learning the performance of Group I was facilitated while that of Group II was hindered ($p < .05$).

Several recent authors (Horowitz, 1962; Underwood, Runquist, & Schulz, 1959) have viewed the paired-associate (PA) method as a combination of two tasks: *S* must learn to produce a set of responses (response learning), and he must also learn to associate each response with a particular stimulus (associative learning). The importance of this analysis was first recognized by Underwood, Runquist, and Schulz (1959), who were studying the effects of intralist similarity on PA learning. Underwood et al. demonstrated that response learning is facilitated early in PA learning if the responses are highly similar, but that interference may occur during the associative phase (Feldman & Underwood, 1957).

A recent study by Gannon and Noble (1961) raises a methodological issue which seems related to this analysis of the PA method. To understand how a familiarization task affects later PA learning, Gannon and Noble familiarized *S* with a list containing the PA stimuli in one condition, the PA responses in another, both stimuli and responses in a third condition, and only irrelevant items in a fourth condition. In all conditions, however, these items were embedded in a longer list containing at least 10 other irrelevant items. In their study stimulus familiarization facilitated PA learning while response familiarization did not, a result contrary to other recent findings (Sheffield, 1946; Underwood & Schulz, 1960; Weiss, 1958).

To help account for this result, Schulz and Tucker (1961) have called attention to a critical feature of Gannon and Noble's procedure: In the PA task *S* was required to pronounce the stimulus before anticipating the response. An *S* familiarized with the stimuli can produce these stimuli more rapidly during PA learning and therefore has a longer anticipation period available. To illustrate the importance of this point, Schulz and Tucker have demonstrated that stimulus familiarization facilitates PA learning only if *S* has to pronounce the stimuli.

---

[1] This research was supported in part by Grant M-5428 of the United States Public Health Service. The article is based upon a thesis submitted by the junior author to Stanford University in partial fulfillment of the requirements for the degree of Master of Arts.

The experiment of Schulz and Tucker, however, does not explain how stimulus or response familiarization can hinder PA learning. An analysis of the PA task into response learning and associative learning suggests one source of interference. Since a familiarization procedure is essentially a response learning task, familiarization may interfere with the response learning phase of PA learning. If so, then this source of interference could be minimized by testing *S*'s performance with some task that does not require him to produce the responses—for example, all the correct responses might be placed before him.

One technique that does not require *S* to produce the responses is the associative matching method (Horowitz, 1962). In this method *S* studies a series of paired items and then receives two types of material: a formboard containing all the stimulus items, and a stack of small cards containing one response item each. He is asked to reconstruct the associations, but he need not produce the responses. This method is a variation of the reconstruction method used by Munsterberg and Bigham (1894), Gamble (1909), and Luh (1922).

If performance is measured by the PA method, though, a response learning component is present in *S*'s task. This response learning component might be impeded if *S* has just learned irrelevant but similar responses during familiarization. In one condition of the Gannon and Noble experiment, for example, *S* was not only familiarized with 10 relevant responses, but he also learned 20 highly similar, irrelevant responses. Thus, two transfer effects might have been confounded: facilitation in the associative phase of PA learning and interference in the response learning phase.

The next step, then, in an analysis of the PA method is to compare the transfer from a response learning task (a familiarization procedure): (*a*) to associative matching, where interference from prior response learning is minimized, and (*b*) to PA learning, where it is not. If the above analysis is correct, a response learning task with only relevant responses will facilitate both associative matching and PA learning. On the other hand, a response learning task with irrelevant, similar responses will impede PA learning but not associative matching. The purpose of the present study is to test this hypothesis.

METHOD

Materials

Japanese words were selected as verbal materials for several reasons. For one thing, Japanese words vary in length. This feature eliminates one severe restriction on nonsense syllables, permits a considerable range of intralist similarity, and provides a highly realistic task for *S*. Second, Japanese words are unfamiliar to most English-speaking *S*s. Thus, they can be conveniently transliterated into the Roman alphabet and still remain relatively free of the restrictions imposed by *S*'s verbal habits. Third, psychological data are already available on the Japanese language from studies of phonetic symbolism. And finally, results based on such material can be applied directly to help understand efficient methods of teaching a foreign language.

Twelve pairs of Japanese words were selected from the list of pairs supplied by Tsuru and Fries (1933). In their list the 2 words of each pair have opposite meanings and could be used to form two lists of 12 words each. Words were selected only if their English translation occurred at least 10 times per million in the Thorndike-Lorge (1944) frequency count.

The items of List I were: SWEET-AMAI, THICK-ATSUI, NEAR-CHIKAI, LOW-HIKUI, HARD-KATAI, BLAME-KENASU, SLOW-OSOI, SMART-RIKO, WHITE-SHIRO, KITE-TAKO, BIRD-TORI, and YOUNG-WAKAI. The items of List II were: DULL-BAKA, FAST-HAYAI, PRAISE-HOMERU, BOAT-HUNE, BITTER-KARAI, BLACK-KURO, WORM-MUSHI, OLD-OITARU, HIGH-TAKAI, FAR-TOOI, THIN-USUI, and SOFT-YAWAI.

Successive trigrams within each Japanese word were examined, and the mean frequency of their occurrence in English was computed. These trigram frequencies were obtained from the norms supplied by Pratt (1939).[2] The mean trigram frequencies of List I ranged from 0.0 to 27.5 with a median of 5.3; in List II they ranged from 0.0 to 28.2 with a median of 5.7.

Experimental Procedure

Each *S* participated first in a practice task and then in one of four response learning tasks. Then *S* attempted either an associative matching task or a PA task. Thus, there were eight different combinations of experimental conditions, which are described below in detail.

Practice trials.—The *S* was presented with 12 CVC trigrams one at a time for 3.5 sec. each on 3 × 5 in. index cards. After each presentation he recorded the items that he could remember. This procedure was repeated to make a total of three trials. Thus, *S* was acquainted with the response learning task, and the experimental groups were equated for their performances.

Response learning.—The *S* was presented with 12 foreign words singly on 3 × 5 in index cards for 3.5 sec. each. He was then given unlimited time to record as many of these items as he could remember. The order of presentation of the items was systematically varied from trial to trial to minimize *S*'s self-instruction for ordering. This procedure was repeated to make a total of five trials.

There were four response learning conditions. The *S*s of Group I were presented with the same Japanese words that would later appear in the associative matching or PA task. Half of these *S*s were presented with List I items, and the remaining *S*s, with List II items. For the *S*s of Group II, the 12 items presented for response learning differed from those later presented in the associative matching or PA task. Half of these *S*s were presented with List I items for response learning and the remaining *S*s with List II items. (The alternate list was then used for associative matching or PA learning.) The *S*s of Group III were presented with a list of Russian words which were transliterated into the Roman alphabet, and the *S*s of Group IV did not participate in any response learning task.

Associative matching or PA learning.—Immediately after the response learning task, half of the *S*s of each response learning condition participated in an associative matching task. In this task *S* was presented with 12 3 × 5 in. index cards singly for 3.5 sec. each. Each card contained one stimulus-response pair. The response of each pair was a Japanese word and the stimulus was its English translation.

After all pairs had been presented once, *S* was supplied with a formboard 8 × 12 in. containing all 12 stimuli. He also received a packet of 12 cards $1\frac{1}{2} \times 2\frac{1}{2}$ in., and each card contained one response. The *S* was instructed to place each response onto the formboard beneath the stimulus with which it had been paired. When he completed this matching, the pairs were presented again, and *S* then repeated the matching task. This procedure was repeated to make a total of five trials. On each trial the order of presentation of pairs was systematically varied, and so was the position of each stimulus item on the formboard. The *S* was given unlimited response time, and his score was the number of items he matched correctly on each trial.

The remaining half of the *S*s in each response learning condition were presented with a PA task immediately after response learning. These *S*s were presented with the same 12 pairs of items described above. After each presentation, 12 cards were presented to *S*. Each card contained one stimulus word, and *S* was given unlimited time to record the corresponding response. This procedure was repeated to make a total of five trials. The order of presentation of stimulus-response pairs was varied systematically, and so was the order of presentation of stimuli during recall.

Subjects

Eighty students of the introductory psychology class at Stanford University served

---

[2] The Underwood and Schulz (1960) norms were not available when the present study began. Since their publication, however, the mean trigram frequencies reported here have been recomputed. According to the Underwood and Schulz norms, the mean trigram frequencies for the items of List I ranged from 0.0 to 361.0 with a median of 23.0; for List II they ranged from 1.0 to 125.0 with a median of 23.6. The rank order correlation coefficient between the mean trigram frequencies computed from the two sets of norms was .94 for List I and .98 for List II.

TABLE 1

MEAN NUMBER OF CORRECT RESPONSES, MODIFICATIONS, AND IMPORTATIONS ON EACH TRIAL OF RESPONSE LEARNING

Trial	List I			List II		
	Correct Responses	Modifications	Importations	Correct Responses	Modifications	Importations
1	2.40	0.80	1.00	2.80	1.20	0.65
2	4.45	0.75	0.55	5.10	1.25	0.45
3	5.55	1.30	0.40	7.10	1.50	0.35
4	5.85	1.20	0.60	7.35	1.85	0.30
5	6.90	0.95	0.65	8.90	1.25	0.10

individually as *S*s for the present study. These *S*s had not been formally trained in Japanese or Russian, nor had they participated in any other verbal learning experiments.

RESULTS

The most general result is that relevant response learning, the task of Group I, facilitated both associative matching and PA learning; similar but irrelevant response learning, the task of Group II, facilitated associative matching and hindered PA learning. The detailed results for each learning task are reported below.

TABLE 2

ANALYSIS OF VARIANCE OF THE SCORES EARNED THROUGHOUT FIVE TRIALS OF ASSOCIATIVE MATCHING

Source	*df*	*MS*	*F*
Between *S*s	39		
Lists (L)	1	0.180	—
Conditions (C)	3	13.633	—
L × C	3	20.673	—
Between *S*s within combinations	32	15.210	
Within *S*s	160		
Trials (T)	4	75.568	48.817***
T × L	4	2.668	—
T × C	12	5.104	3.297***
T × L × C	12	1.911	—
T × *S*s within combinations	128	1.548	
Total	199		

*** $p < .001$.

Correct Responses

Response learning.—Table 1 presents the total number of correct responses on each trial of response learning for each list of Japanese words. The scores for List II were uniformly higher throughout the five trials, and this superiority increased on successive trials. An analysis of variance of the correct responses showed that the variation between lists was significant at $p < .05$ ($F = 4.45$). The List × Trials interaction was also significant at $p < .05$ ($F = 3.00$).

Associative matching.—The analysis of variance of the correct matching scores is summarized in Table 2. The between-list difference was not statistically significant, nor was its interaction with experimental conditions or with trials. Whatever characteristic made List II items easier to recall in response learning did not seem to influence associative learning. Therefore, in subsequent analyses of the associative matching scores, the data from the two lists have been combined.

Although the average difference between experimental conditions was not statistically significant, the interaction between conditions and trials was. The presence of this interaction is apparent from Fig. 1, which reports the mean correct matching score on each trial for each of the experimental conditions. Initially, the performance of Groups I and II, whose response learning task contained Japanese words, was superior to that of Groups III and IV, the control groups. This difference diminished as the groups approached asymptotic performance.

Thus, response learning facilitated associative matching when the responses learned were at least similar to those of the associative matching task. When dissimilar (Russian) responses were learned, though, no such transfer occurred. These results suggest that response learning facilitates associative matching only if the response learning task enhances the discriminability of the associative matching items. In the present study, for example, *S* must have learned some of the characteristic redundancy in Japanese words and could thus attend more readily to the distinctive features of the items during associative matching.

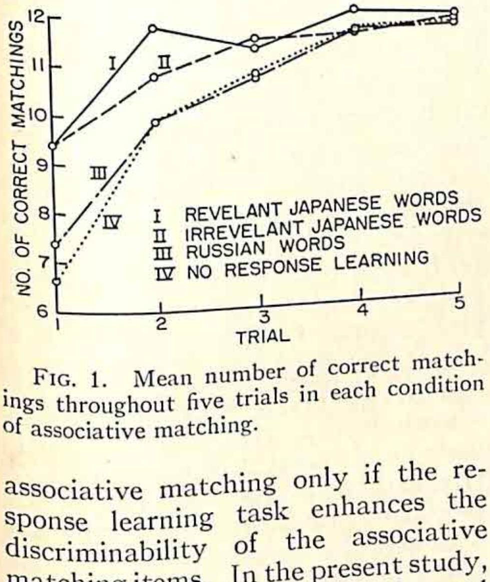


FIG. 1. Mean number of correct matchings throughout five trials in each condition of associative matching.

PA learning.—Figure 2 reports the mean number of correct responses on each trial of PA learning for each experimental condition. An analysis of variance was performed on these data, with the results summarized in Table 3. These results, like those of the associative matching task, show a significant interaction between trials and conditions because the difference in performance among the four groups decreased on successive trials. The major difference between these results and those of the associative matching task is in the performance of Group II. Although the response learning task

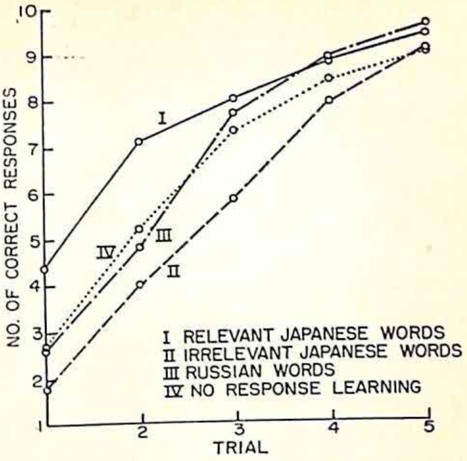


FIG. 2. Mean number of correct responses throughout five trials in each condition of PA learning.

TABLE 3

ANALYSIS OF VARIANCE OF THE SCORES EARNED THROUGHOUT FIVE TRIALS OF PA LEARNING

Source	*df*	*MS*	*F*
Between *S*s	39		
Lists (L)	1	75.645	3.008
Conditions (C)	3	28.498	—
L × C	3	16.018	—
Between *S*s within combinations	32	25.150	
Within *S*s	160		
Trials (T)	4	275.668	141.151***
T × L	4	1.558	—
T × C	12	3.828	1.960*
T × L × C	12	0.364	—
T × *S*s within combinations	128	1.953	
Total	199		

* $p < .05$.
*** $p < .001$.

of Group II facilitated associative matching, it hindered PA learning.

Analysis of Errors

Response learning.—Table 1 also presents the total number of errors during response learning categorized into "modifications" and "importations." A modification is defined as a response with only one letter faulty (i.e., omitted, added, incorrect, or misplaced). All other errors were called importations. The proportion of errors in the responses of List I did not differ significantly from that of List II (.25 for List I and .22 for List II), but the types of errors differed markedly. Of the errors of List I, .61 were modifications; of those of List II, .79 were modifications. The difference between these two proportions was significant at $p < .001$.

Associative matching.—The only possible type of error in an associative matching task is an intralist confusion between two or more responses. Matrices were compiled to show how often each response was mistakenly substituted for each correct response, but the absolute numbers of such confusions were too small to permit any generalizations. The confusion matrices of Groups I and II were pooled, and those of Groups III and IV were pooled. The marginal totals of the pooled matrices gave frequency distributions reporting how often each item was missed. These two distributions differed markedly. The items ATSUI, CHIKAI, HIKUI, and WAKAI accounted for 41% of the 183 errors of Groups III and IV, but for only 24% of the 103 errors of Groups I and II. The value of r computed between these frequency distributions was only .46, which is just barely significant at $p < .05$. Thus, a response learning task with relevant or irrelevant Japanese items seems to have had two effects: It reduced the total number of errors, and (perhaps because) it changed the items that induced errors most often.

TABLE 4

PERCENTAGE OF ERRORS OCCURRING IN EACH CATEGORY THROUGHOUT FIVE TRIALS OF PA LEARNING

Category of Error	Group I (%)	Group II (%)	Groups III & IV (%)
1. Errors of associative learning only	14.0	4.8	6.5
2. Modifications with correct associative learning	20.7	19.7	19.3
3. Modifications with faulty associative learning	7.7	7.0	7.3
4. Between-list intrusions	0.0	0.9	0.0
5. Other intrusions	14.4	17.1	20.0
6. Failures to respond	43.2	50.5	46.9

PA learning.—The errors during PA learning were classified into six categories, as reported in Table 4. Category 1 contained responses which were correct but were elicited by the wrong stimulus; i.e., errors of associative learning. Categories 2 and 3 contained modifications, as defined above. A modification was placed in Category 2 if it was elicited by the correct stimulus (error of response learning only) but in Category 3 if it was elicited by an incorrect stimulus (errors of both response learning and associative learning). Category 4 contained the correct responses and modifications that were appropriate to the alternate list. Such errors were directly due to interference from the similar, irrelevant response learning task. Category 5 contained all other intrusions, and Category 6 contained the failures to respond.

The mean number of errors made by Group I throughout five trials was 22.2 ($SD = 10.1$); by Group II, 31.5 ($SD = 7.7$); by Group III, 25.8 ($SD = 6.8$); and by Group IV, 27.7

(*SD* = 12.3). The nature of the errors made by Groups III and IV were so similar that they have been pooled to give the percentages reported in Table 4. The *S*s of Group I made proportionately more associative matching errors, and the *S*s of Groups III and IV made the most severe response learning errors. Surprisingly, the errors of Group II were almost never between-list confusions; instead, these *S*s produced fewer responses.

DISCUSSION

The present authors cannot account for the difference between Lists I and II in the number of correct responses during response learning or in the kinds of errors produced. At first, the difference seemed to be due to a difference in intralist similarity between the lists. A fact supporting this explanation was that the mean trigram frequencies (computed from the Underwood and Schulz norms) were more variable for List I items than for List II items. (The *SD* of the mean trigram frequencies of List I was 96.3; that of List II, 44.7.) If greater variability among mean trigram frequencies contributes greater discriminability to a list of items, the intralist similarity of List II is greater, and response learning on it should be superior (Horowitz, 1961). However, in view of the interaction between the nature of *S*'s task and intralist similarity (Horowitz, 1962), this interpretation would be valid only if performance on associative matching were superior for List I. The two lists did not differ in associative matching, though.

These data have shown that a response learning task can facilitate associative matching when it helps make the associative matching items more discriminable. However, it can interfere with the response learning phase of PA learning when the responses learned are similar to the PA responses. (This result implies that stimulus familiarization might hinder PA learning if the familiarized stimuli are sufficiently similar to the PA responses.) To demonstrate the facilitative effect of response familiarization, *E*s have usually familiarized *S* with only relevant responses. Gannon and Noble, however, produced negative transfer by familiarizing *S* with irrelevant items, as well as relevant ones.

Another variable that probably influences the nature of the transfer in a PA task is the relative difficulties of associative learning and response learning. (Some evidence in the present study suggests that the associative phase of the PA task was easier than the response learning phase. For one thing, the associative matching task was considerably easier than the response learning task. Furthermore, the control groups made fewer associative errors in PA learning than Group I, though the groups were performing equally well by Trial 5. To achieve this high performance level by Trial 5, the control groups must have mastered associative learning while they were still emitting the response learning errors of Category 5.) If associative learning in PA tasks is necessarily easier than response learning, then any facilitation in associative learning alone would contribute relatively less to the net transfer effect; and additional interference in the response learning phase would obscure this facilitation. On the other hand, if PA tasks can be found where associative learning is more difficult than response learning, then the present results may have limited generality, and the net effect of irrelevant response learning might be facilitation.

REFERENCES

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53,** 11–15.

GAMBLE, E. A. A study in memorising various materials by the reconstruction method. *Psychol. Rev. monogr. Suppl.*, 1909, **10** (4, Whole No. 43).

GANNON, D. R., & NOBLE, C. E. Familiarization (*n*) as a stimulus factor in paired-associate verbal learning. *J. exp. Psychol.*, 1961, **62,** 14–23.

HOROWITZ, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.*, 1961, **62**, 51–57.

HOROWITZ, L. M. Associative matching and intralist similarity. *Psychol. Rep.*, 1962, **10**, 751–757.

LUH, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31** (3, Whole No. 142).

MUNSTERBERG, H., & BIGHAM, J. Memory. *Psychol. Rev.*, 1894, **1**, 34–38, 453–61.

PRATT, F. *Secret and urgent.* Indianapolis: Bobbs-Merrill, 1939.

SCHULZ, R., & TUCKER, I. F. Stimulus familiarization in paired-associate learning. Paper presented at Psychonomic Society, New York, September 1961.

SHEFFIELD, F. D. The role of meaningfulness of stimulus and response in verbal learning. Unpublished doctoral dissertation, Yale University, 1946.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* Teachers College, Columbia University, Bureau of Publications, 1944.

TSURU, S., & FRIES, H. S. A problem in meaning. *J. gen. Psychol.*, 1933, **8**, 281–284.

UNDERWOOD, B. J., RUNQUIST, W. M., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WEISS, R. L. Role of association value and experimentally produced familiarity in paired-associate learning. Unpublished doctoral dissertation, University of Buffalo, 1958.

(Received March 7, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 233–239

CAN THE ABSOLUTE THRESHOLD BE CONDITIONED?[1]

T. S. AIBA[2]

Institute of Psychiatry, University of London

Earlier studies have suggested that the absolute threshold to light can be raised by certain irrelevant stimuli, such as a tone, that had previously been presented many times in the presence of adapting lights. An attempt was made to confirm this finding. 2 groups took part, each consisting of 6 Ss. The 1st group received a combination of a tone and a moderately bright light, the 2nd a tone and a weak red light. Later, the threshold was measured in the presence of the tone alone. Of the 2 groups, the 1st gave consistently higher thresholds. The effect is not attributable to adaptation. It is concluded that at least the possibility of the conditioning of the absolute threshold exists.

In their recent article, Nebylitsyn, Rozhdestvenskaya, and Teplov (1960) described a method involving conditioning which purports to give a measure of certain sensory-behavioral characteristics of human *S*s.

In spite of its broader implications, their study may be thought of as belonging to the realm of sensory investigations, since their particular choice of type of conditioning involved a conditioned decrease of the absolute threshold to light. As the issue involved is of some importance to visual psychophysics, it seemed appropriate to make a critical appraisal of the methods they used and then to repeat part of their experiments to see whether or not the conditioning effect they obtained was real.

Nebylitsyn et al. caused a transitory rise in the thresholds of the dark-adapted eye of their *S*s by presenting to them an illuminated panel for a few seconds (i.e., light adaptation), and at the same time giving them a pure tone (or weak red light), which also lasted for a few seconds. They repeated this pairing of the light-adapting field and the tone (or the weak red light) a number of times, and eventually found that the dark-adapted threshold of the *S*s could be raised merely by presentation of the tone or the weak red light alone.

The principal question concerns the psychophysical method used by Nebylitsyn et al. for determining the absolute thresholds (which they call "the background sensitivity") and the other transient thresholds which were caused by the light-adapting stimuli and by what they called their conditioned stimuli (i.e., the tone or the weak red light). They used a sort of methods of limits wherein they first presented to the *S*s the test target of an adaptometer either supra- or subliminally, and then quickly changed the intensity until the target either vanished or appeared. The *S*s were required to report the disappearance or the emergence of the stimuli as quickly as possible. This method not only makes the state of adaptation of the *S*'s eye ambiguous in that the target was exposed for unspecified periods of time, but also makes such extraneous factors as the *S*'s RT and the speed at which the *E*s changed the intensity of the test stimulus too important in the threshold measurement. This criticism applies also to the

---

[1] Supported by a grant from the Human Ecology Fund.

[2] Now at Psycho-Acoustic Laboratory, Harvard University.

methods used by Rozhdestvenskaya (1955).

There are other details of their work that may also be questioned, but these concern more the way in which the experimental procedures are reported in the two papers. For example, both Nebylitsyn et al. and Rozhdestvenskaya neglect to specify the luminances of their unconditioned stimuli (light-adapting fields) in any of the accepted photometric units—they give certain values of illumination only. In addition, both papers fail to report the absolute values of the thresholds, and give *percent decrease or increase in sensitivity*. Probably a more serious omission in their reports is their failure to give the variability of their background thresholds, which must have been an important clue to whether a given threshold rise following the tone or the weak red light was really a conditioned response, or just a chance variation. Presumably, the variabilities they encountered in their experiments were rather small, because as their criterion for a successful conditioning they used three successive occurrences of the conditioned threshold-rise which amounted to only 1 db. on the average.[3] (In Rozhdestvenskaya's study, the threshold rise was only 0.8 db. on the average when the pupil size was controlled.)

It is of interest to see whether the Russian investigators' results could be duplicated if better control were exerted over such variables as are mentioned above, and if the use of experimental and control groups eliminated the possibility that the contioning is an artifact inherent in the particular procedures used. Such an experiment has been conducted, and the results obtained are shown below.

METHOD

Apparatus

The optical part of the apparatus consisted essentially of three optical systems, each one making use of the Maxwellian viewing arrangements. At the end of one of the systems was placed an adaptometer. The original size of the test target in the adaptometer had been reduced so that the light source was just large enough for the Maxwellian system but emitted no unwanted stray light around the apparatus. A motor-driven shutter, attached to the adaptometer just over the test target, exposed the target for 2 sec. every 5 sec. The light from the test target passed through a pair of neutral density wedges and a set of lenses, and was focused on the plane of an artificial pupil, 3 mm. in diameter, placed at the other end (on *S*'s side) of the system. The light source for the second optical system was a 12-v. motor-car side-lamp, housed in a metal box which was light-tight except for a small circular aperture covered by a translucent glass. The light from this glass was, after being collimated by a lens, reflected on a thin glass plate and brought into the first optical system to be focused on the plane of the above-mentioned artificial pupil. The third system, which provided the fixation point, consisted of a lens and a light source similar to the second system but covered by a deep red filter. The third system was also brought into the first system by means of a plain glass plate reflecting the light, and the image of the source was focused on the artificial pupil. The intensity of the source in the first system was altered by varying the diaphragm setting of the adaptometer, as well as by changing the relative positions of the wedges. The wedges, which were mounted on rack and pinion mechanisms, could be moved either continuously or in steps of 0.45 db. In either case, the density of the wedges could be read from millimeter scales attached above them.

The intensity of the source in the second system was regulated by the ac current which lit the lamp, and its luminance was adjusted to an adequate level by a set of neutral density filters put in the path of the light. There was also a provision for placing a color filter in front of the source. The luminance of the fixation spot, which was to be maintained at a just visible level during the experiment, was controlled by a rheostat connected to the lamp.

---

[3] The decibel is a logarithmic measure of relative energy or power. The number of db. equals 10 log A/A_0. Absolute levels of luminance are conveniently expressed by setting A_0 at 10^{-10} L. For advantages in using db. in vision, see Stevens (1955, 1957).

The *S* looked at the stimuli through the 3-mm. artificial pupil, and his position was fixed by means of a biting board and a headrest.

A transistorized oscillator of the multivibrator type produced a pure tone of 1000 cps. The durations of this tone, as well as of the light in the second optical system, which served as the CS and the UCS, respectively, were controlled by electric relay switches operated by a uniselector. The uniselector was activated by pulses generated in the shutter mechanism controlling the exposure of the test target in the adaptometer.

Procedure

Two groups of *S*s took part in this study—an experimental and a control group. The two groups were given exactly the same instructions and as far as the *S*s were concerned performed exactly the same tasks. The only difference was that during the conditioning sessions, they were given different kinds of light for their unconditioned stimuli. The following description of the experimental procedures applies to both the experimental and control groups, unless otherwise specified.

Threshold measurements.—All threshold measurements were of test flashes, lasting .2 sec., subtending a visual angle of 3° and centered on a spot 7° away from the fixation point. At the first two sessions, on 2 separate days, preliminary measurements were made of the absolute thresholds by the usual method of limits. The *S*s were also familiarized with the experimental situations, which included presenting the tone a number of times to eliminate the orienting reactions (see Nebylitsyn et al., 1960, p. 19). In the third session, which began the actual experiment, *S* was first dark-adapted for 40 min., and the intensity of the test flashes was then determined at which *S* reported seeing light approximately 60% of the time. The reason for using this particular frequency of seeing was as follows: on the usual frequency-of-seeing curve, 60% of seeing would fall at the most linear and steepest portion of the curve. Thus if *S*'s threshold increased or decreased slightly, this change would be reflected most sensitively in a corresponding decrease or increase in the frequency of seeing. The present study was more concerned with a decrease than an increase in sensitivity. Thus the 60% point was considered more suitable than other points in this region, say, the 40% point, in that it would allow the subsequent decrease in the frequency of seeing, without in any way altering the relation between the sensitivity and the frequency of seeing mentioned above.

To determine the 60% "seen" intensity, four or five levels of intensity likely to give this frequency of seeing had been selected in advance on the basis of *S*'s previous threshold measurements. These intensity levels were in steps of .9 db., and the test flashes were presented 20 to 30 times at each of these levels in a random order in accordance with the method of constant stimuli. There were presented also about an equal number of blanks, randomly mixed with the actual flashes, in which *S* heard the click of the shutter but no light was presented. This procedure was taken in an effort to ensure that *S*'s reports were not arbitrary. When *E* found too many false positive responses to the blanks (if more than 10% of the blanks were reported as "seen"), *S* was told of this fact and encouraged to do better. Most *S*s' performances improved after such a comment from *E*, but if not, the experiment was discontinued and the *S* released.

Conditioning, or pairing of the light and the tone.—After the *S*'s initial threshold measurement was completed, conditioning was begun. The procedure was as follows: the 1000-cps tone was presented to *S* through a pair of earphones for 15 sec., and while the *S* was listening to this tone, a white light of 72 db. subtending a visual angle of 10° was presented to the experimental group for 10 sec., and a dim deep red light, which actually consisted of the same white light transmitted through an Ilford 609 deep spectrum red filter and a 6-db. neutral density filter, was presented to the control group for 10 sec. Both these lights were centered on a spot where the test flash normally appeared. They were both clearly visible to *S*s. The intensity of the tone was around 70 db. above the threshold; i.e., the level at which *S* could hear the tone clearly but not high enough to cause discomfort. The lights were presented 5 sec. after the onset of the tone, but they terminated at the same time as the tone in accordance with the procedure used in the study of Nebylitsyn et al.

The pairing of the tone and the light was repeated 12 times during a single day's session. Intermingled randomly with these were eight trials, each consisting of the presentation of the conditioning tone (1000 cps, 15 sec.) followed 5 sec. after its termination by 10 test flashes at the intensity that had been determined to give 60% frequency

TABLE 1

NUMBER OF FLASHES SEEN IN EACH TRIAL AND THRESHOLD (60% SEEN) AT THE BEGINNING OF EACH SESSION

Subject	Session	Threshold (in db.)[a]	Trials 1	2	3	4	5	6	7	8
Control Group										
VA	I	50.9	5	3	2	3	5	7	5	6
	II	49.5	4	9	9	9	5	8	8	10
	III	50.2	9	7	9	5	10	9	9	10
RR	I	53.6	2	2	1	9	4	10	5	1
	II	54.9	9	6	8	8	7	7	9	9
	III	54.9	3	7	6	7	9	5	6	5
MS	I	50.9	6	8	9	5	4	2	4	7
	II	50.0	6	4	5	4	7	10	7	9
	III	52.2	0	6	2	8	9	5	7	4
SS	I	50.4	10	10	9	6	9	5	9	9
	II	51.3	8	1	0	2	0	3	2	5
	III	50.4	7	9	6	8	9	2	3	6
JL	I	52.6	6	4	5	6	7	4	7	2
	II	51.8	7	3	3	0	7	9	3	6
	III	53.1	4	1	2	5	2	7	8	1
ES	I	50.4	9	9	7	5	4	6	7	10
	II	50.4	5	7	6	7	4	8	8	5
	III	51.3	6	2	5	4	8	6	9	5
Experimental Group										
LS	I	54.0	5	5	6	4	3	1	5	5
	II	53.3	4	2	6	1	0	8	4	3
	III	52.2	9	9	4	4	7	4	9	5
TF	I	52.7	5	7	3	4	8	6	3	7
	II	51.9	4	1	3	3	2	1	5	0
	III	51.3	3	1	3	6	3	1	1	2
MW	I	50.9	3	4	6	5	1	3	2	2
	II	50.0	7	6	9	3	6	4	3	4
	III	53.1	3	6	1	3	1	3	3	1
GF	I	53.1	6	4	5	6	6	5	7	1
	II	52.2	5	3	1	2	1	5	2	0
	III	53.1	1	5	1	6	1	7	2	4
CP	I	53.1	4	8	7	7	3	6	2	4
	II	51.3	8	6	5	5	7	3	4	5
	III	53.1	3	9	8	0	10	0	1	1
AI	I	50.4	5	7	7	7	5	10	5	6
	II	52.2	3	8	6	3	3	1	1	1
	III	51.8	0	8	9	2	6	5	2	0

[a] The decibel values are in reference to 10^{-10} L., 50 db. corresponding to −2 log mL.

of seeing. The test flashes were delivered at the rate of one in every 5 sec. The *S* was asked to report whether or not he had seen light after each presentation of the test flash.

When a trial followed a pairing of tone and light, at least 6 min. was allowed for the eye to recover fully from the light adaptation before the trial was begun. This was done with the control group as well, in order to equate conditions. The 6-min. interval was considered sufficient in view of the results obtained in the preliminary experiments. Similarly, when one pairing followed another, at least 3 min. was allowed between them, with the result that the unconditioned stimuli in the two sessions were both uninfluenced by the light adaptation.

The 12 pairings and eight trials were repeated again, in a slightly different order, in the two subsequent sessions on different days. Each session lasted about 2 hr. and 20 min., including the period for dark adaptation.

Subjects

All *S*s were paid volunteers from all walks of life. None of them knew the true purpose of the experiment or any facts concerning the visual processes in light and dark adaptation. All had normal vision. There were 8 women and 4 men. Their ages were between 19 and 30, with a mean of 23.5.

RESULTS

Table 1 shows the number of flashes seen by each *S* in each of 24 trials during the three sessions. It also gives the individual *S*'s thresholds, as defined by the intensity of the test flashes required for 60% frequency of seeing.

Figure 1 is a graphic presentation of the data in Table 1, together with *SD*s showing variabilities among the *S*s. There were, on the average, 1.5 pairings of the light and the tone between each pair of trials, and the first trial was also preceded by an average of 1.5 pairings. Consequently the number of trials multiplied by 1.5 gives the number of pairings completed on that day.

The most prominent features of the results are as follows: First, the control and experimental groups do not differ very much at the beginning of the session, but as more and more pairings are added the difference begins to be more apparent. Furthermore, this tendency is almost the same for all 3 experimental days (Sessions I, II, and III).

Second, the difference between the two groups in the latter half of the session (Trials 5–8) appears to be greater in Session II and Session III than in Session I. Between Sessions II and III this difference is not so apparent.

Third, the control group more or less retains its original level of seeing throughout any of the sessions (i.e.,

TABLE 2

Scores Based on the Four Summed Trials

Control Group				Experimental Group			
Subject	Session	Trials[a] 1	Trials[a] 2	Subject	Session	Trials[a] 1	Trials[a] 2
VA	I II III	13 31 30	23 31 38	LS	I II III	20 13 26	14 15 25
RR	I II III	14 31 23	20 32 25	TF	I II III	19 11 13	24 8 7
MS	I II III	28 19 16	17 33 25	MW	I II III	18 25 13	8 17 8
SS	I II III	35 11 30	32 10 20	GF	I II III	21 11 13	19 8 14
JE	I II III	21 13 12	20 25 18	CP	I II III	26 24 20	9 19 12
ES	I II III	30 25 17	22 25 28	AL	I II III	26 20 19	26 6 13

[a] Trial 1 refers to the sum of the four trials in the earlier half of the session, Trial 2 in the later half of the session.

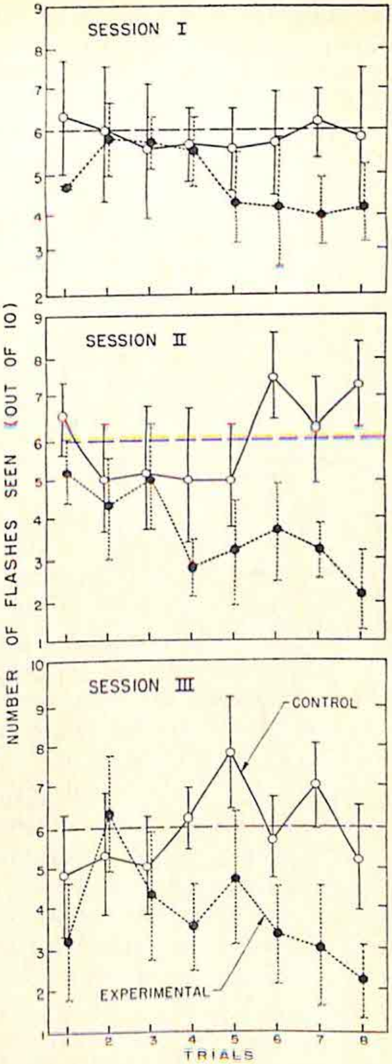


Fig. 1. Number of flashes seen in each of the trials in Sessions I–III. (Each point is the average of 6 *Ss*. The vertical bars show the variability among the *Ss* by *SD*. The horizontal line in each figure indicates the original level of seeing, prior to conditioning, in that session. There were, on the average, 1.5 pairings of the light and the tone between each pair of trials; and the first trial was also preceded by an average of 1.5 pairings. Consequently the number of trials multiplied by 1.5 gives the number of pairings completed on that day.)

TABLE 3

OUTCOME OF ANALYSIS OF VARIANCE FOR DATA IN WHICH FOUR TRIALS WERE SUMMED

Source	*df*	*MS*	*F*
Treatments (T)	1	925.0	19.8[a]**
People within treatment	10	46.6	1.04
Trials (R)	1	18.0	0.40
Sessions (S)	2	29.5	0.66
T × S	2	41.3	0.93
T × R	1	26.5	0.59
T × S × R	2	192.4	4.3*
Residual	52	44.6	
Total	71		

[a] This was compared with the *MS* for people; all others were compared with the *MS* for residual variance.
* $p < .05$.
** $p < .01$.

maintains the 60% frequency of seeing; see the broken lines in Fig. 1), whereas the experimental group tends to fall below this level. Thus the divergence between the two groups, pointed out above, can be regarded as due to the experimental group's decline, rather than the control group's improvement, in the level of seeing.

Now in order to test the statistical significance of the above results, an analysis of variance was carried out. For this purpose, the records of four consecutive trials were summed so that each single session of any *S* is now represented by two scores, one for the first half and another for the second half of the session. This was done in order to have a better comparison between the early and later trials in a session and also to prevent the boosting of the degrees of freedom. The new set of scores, with four trials summed, is given in Table 2, and the outcome of the analysis of variance in Table 3. As can be seen from Table 3, Treatments (i.e., conditioning) and the interaction among Trials, Sessions, and Treatments came out statistically significant. This indicates that the conditioning effect, manifested by the differences between the two groups, is dependent not only on the treatment per se, but also on the number of times the treatment was repeated.

In order to have a clearer idea of how the conditioning effect was formed during the course of the experiment, the differences between the two groups were analyzed using the differences between the means of the two groups in all periods of the experiment; i.e., in the earlier and later trials of each of the three sessions. The device used was the *t* test for uncorrelated means using the residual variance obtained in the analysis of variance as an index of variance. It turned out that only in the later trials of Sessions II and III did the two groups significantly differ; the *t*'s for the differences between the two groups in the earlier trials in Sessions I, II, III were 0.467, 1.141, and 1.037, respectively, while the differences in the corresponding values of *t*'s for the later trials were 1.686, 3.579, and 3.242, the last two being significant beyond the 2% level of significance.

DISCUSSION

It may be that the observed effects constitute conditioning. Since the criterion for successful conditioning was statistically significant differences between the experimental and control groups, the conditioning effects obtained cannot be considered as artifacts, for fatigue or chance variations in thresholds should be operating to the same extent in both groups. Nor can the observed differences between the two groups be due to cumulative light adaption in the experimental group, since preliminary experiment showed that 6 min., the actual interval used, was sufficient for the eye to recover fully from adaptation to a 72-db. light for 10 sec. The fact that the two groups differed, not on the first day, but only on the second and third days, also indicates that light adaptation cannot be responsible. Nor does expectation or set explain these results, since the *S*s of both groups were ignorant of the purpose of the experiment and the effect of white light on their thresholds.

Against these arguments, one may list reasons why the observed effects may *not* be conditioning. For instance, the repeated presentation of a moderately bright white light to the experimental groups may have caused them to raise their criterion for reporting the test flashes, whereas to the control group the UCS (the dim red light) and the test flashes were sufficiently different in appearance not to cause this change. Unlike light adaptation, the change of criterion could develop slowly over a few days. The fact that the conditioning effects could not be maintained from one day to another may also raise some question of whether this was indeed conditioning.

Regardless of the strength of either argument, there would be one point of agreement, that is, that the effects obtained here are extremely small. Even the largest difference between the control and experimental groups in frequency of seeing (i.e., number of flashes seen out of 10) in corresponding trials is no more than 50%. This corresponds to a difference of about 3 db. in the thresholds if the usual frequency-of-seeing curve is considered. All other differences are of the order of 1 db. or less. Such a small effect may be produced by almost anything, by a change in subjective criterion, by past experiences, or by any of the kinds of events that are disturbing to *S*s, and possibly by conditioning, too. Thus, unless more substantial effects are obtained, nothing conclusive can be said on whether the absolute threshold can be conditioned, but it may be said that the results of the present study at least indicate the possibility of such conditioning.

REFERENCES

NEBYLITSYN, V. D., ROZHDESTVENSKAYA, V. I., & TEPLOV, B. M. Concerning the interrelation between absolute sensitivity and the strength of the nervous system. *Quart. J. exp. Psychol.*, 1960, **12**, 17–25.

ROZHDESTVENSKAYA, V. I. The role of pupil in conditioned reflex changes of the sensitivity of the eye in peripheral vision. *Probl. fiziol. Opt.*, 1955, **11**, 25–29.

STEVENS, S. S. Decibels of light and sound. *Physics To-day*, 1955, **8**, 12–17.

STEVENS, S. S. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

(Received April 9, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 240–247

PROBABILISTIC HIERARCHIES TO AMBIGUOUS CONCEPT CLASSES[1]

SOLOMON E. FELDMAN[2]

Indiana University

In a test of a statistical model of visual recognition, 160 Ss learned to anticipate the nonsense syllable names of 4 types of stimulus displays, each of which included an outline figure and a word. Subsequent responses to differing cue combinations indicated that: (a) the probabilities of sampling various cues from among a larger set of cues and the probability of responses to cues that have been sampled are both functions of the relative and absolute frequencies of previous reinforcement, (b) stimulus elements are associated with patterns of cues rather than with the individual cues comprising the patterns, and (c) response dependencies only occur when repeated responses to the same type of stimulus are required.

Binder and Feldman (1960) demonstrated that the relative frequencies of *S*'s responses to ambiguous cues tend to equal the relative frequencies of previous responses to these cues which correctly assigned the cues, as part of a larger set of cues, to unique classes (matching). The present experiment was designed as an attempt to test the range of applicability of this statistical model of visual recognition.

Specifically, the present study was aimed at testing the following hypotheses: (*a*) probability matching will occur despite the fact that class specifying attributes are "abstracted elements" rather than physically identical elements that are constant from trial to trial; (*b*) probability matching will occur whether ambiguity results from having too few cues, having too many cues, or having both too few and too many cues for unique class specification; (*c*) a model which considers stimulus elements to be associated with patterns of cues qua patterns produces better predictions of responses to ambiguous cues then a model assuming association of elements with the individual cues comprising the patterns, and, as a major extension of the previous model; (*d*) the relative frequencies of selective sampling the stimulus elements associated with various ambiguous cues tend to equal the relative frequencies of presentation of these cues when unique class specification was possible.

METHOD

Subjects.—The *S*s, students from introductory psychology sections at Indiana University, were run individually. The difficulty of the stringent learning criterion required the use of 270 *S*s in order to obtain the required 160 *S*s, reaching the learning criterion.

Apparatus.—The apparatus was described fully by Binder and Feldman (1960). Briefly, *S*s sat in a semidarkened room in front of a panel with two one-way vision windows. The length and sequence of the illumination behind these windows, and thus the exposure durations, were under experimental control.

Presentation material.—During the learning phase *S*s were required to anticipate the nonsense syllable name of four different types of stimulus presentations. Each of these presentations consisted of a verb or a noun and a straight-line or curved outline figure printed on the same stimulus card. Examples of the four types of learning stimuli may be seen in Fig. 1.

---

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements of the PhD degree at Indiana University where the author received support from Grant 2M-5425-C9 of the United States Public Health Service. The author wishes to express his appreciation to Arnold Binder, Chairman of the Research Committee, and W. K. Estes, A. M. Buchwald, and R. Milisen who also served on the Committee.

[2] Now at the University of Wisconsin.

A pool of 244 English words one-half of which were "clearly" nouns and one-half "clearly" verbs was obtained by having a separate sample of 100 *S*s classify a larger list of words the *E* previously had selected from the Thorndike and Lorge (1944) compilation. Ten verbs and 10 nouns were randomly chosen with restrictions that permitted controls for length, frequency, and positional preferences for use in the construction of test stimuli.

Following procedures suggested by Attneave and Arnoult (1956) eight points were randomly placed on each of 224 poster board cards. On half of these cards the points were connected using a straight edge, while on the remaining half a protractor having a 2-in. radius was employed.

The point of insertion of a word within a figure was randomly determined as was the particular word employed with restrictions that permitted a counterbalancing of word length and frequency of occurrence in English discourse for the various types of presentations.

Following similar procedures eight types of test presentations were constructed. Four types, consisting of a noun (N), a verb (V), a straight outline figure (S), and curved outline figure (C) presented singly, had too few cues for unique class specification. Two types, a straight figure enclosing both a noun and a verb (SNV) and a curved figure enclosing both a noun and a verb (CNV) had too many cues for unique identification. A noun and a verb with no enclosing figure (NV) presented a situation with both too few and too many cues. The final type (N/V) consisted of words presented singly which were "intrinsically" ambiguous in that the 100 raters were approximately equally divided in their classification of the word.

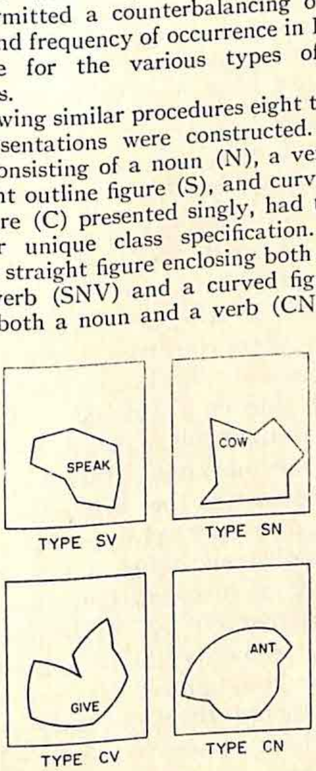


FIG. 1. Examples of the four types of stimulus cards.

TABLE 1

FREQUENCIES OF OCCURRENCE OF THE LEARNING STIMULI/15-TRIAL BLOCK

Learning Stimuli	Group							
	A	B	C	D	E	F	G	H
CV	4	1	8	2	2	1	8	4
SV	1	4	2	8	1	2	4	8
CN	8	2	4	1	8	4	2	1
SN	2	8	1	4	4	8	1	2

Procedure.—Each stimulus card can be considered as a member of a concept class for which *S*s had to learn the appropriate name, e.g., CV as a WOY. The particular nonsense syllable from among WOY, RUX, TYZ, and JOM (each of 33% association value; Glaze, 1928) paired with a given concept class was randomly determined over all *S*s.

Any individual cue (i.e., C, V, N, S) was associated with two responses during learning. Two relative frequencies of association of the cues with their appropriate responses were employed: 2 to 1 and 4 to 1. A counterbalancing procedure was used to control for biases for a particular cue or set of cues. The resulting conditions can be considered as forming eight groups of 20 *S*s each, which differed in the absolute frequency of presentation of the various types of learning stimuli. An *S* was assigned to a particular group according to a random schedule or group assignment. The counterbalanced design and the number of times each word-figure combination was shown for each group can be seen in Table 1.

For the sake of exposition, CV will be assumed as being associated with R_1, SV with R_2, and so on. Looking at each cue singly, it should be noted that although the relative frequency of association of a cue with its two appropriate responses was always either 2 to 1 or 4 to 1, the absolute frequencies of association per trial block differed in the various groups. Thus for both the *S*s in Groups A and B a curved figure was associated twice as often with R_3 as it was with R_1. However, for the *S*s of Group A, the absolute frequency of concept-class name associations per learning block was 8 to 4 while for *S*s of Group B this value was 2 to 1. The counterbalancing procedure permitted an analysis of the effects of both the relative and the absolute frequencies of association.

For any *S* the particular stimulus cards employed were randomly selected from the available pool of 56 cards of a concept class. Order of presentation was randomized over each 15-trial learning block with the restriction that no concept class would appear more than three times in succession.

A paired-associate anticipation procedure was used. During the instructional period the apparatus was demonstrated, the procedure explained, *E* pronounced each of the four nonsense syllables, and the *S* was told he would have to pay attention to both the word and the figure on each trial. One second after the onset of a .5-sec. warning buzzer one of the four concept classes appeared in the left window. This card was exposed for 5 sec. The response card, the appropriate nonsense syllable, appeared in the right window during the last 2 sec. of the stimulus exposure. Both stimulus and response presentations terminated together. The sequence was repeated 4.5 sec. later when the buzzer signaled the start of a new trial.

The *S*s were required to correct all errors or omissions by reading the correct name when it was shown. Learning proceeded until a complete 15-trial block was correctly anticipated. The available pool of stimulus cards permitted a maximum of seven learning blocks in which the criterion could be met.

Following attainment of the criterion, *S* was shown 12 test presentations which were ambiguous in terms of the previous learning. In determining the 12 test presentations for a given *S* two test cards from each of the available pools of Type C, Type S, Type N, and Type V were randomly chosen with the restriction that each available test card would appear an approximately equal number of times for each of the eight groups of *S*s. For a given *S*, with the same restriction, one test card was randomly selected from each of the available pools of Type CNV, Type SNV, Type NV, and Type N/V.

Within the first four test trials one Type CNV, one Type SNV, one Type C or Type S, and one Type N or Type V was presented in randomly determined order with the restriction that each test type appeared equally often as the first test presentation for each of the eight groups. The order of presentation of the remaining eight test cards was randomly determined except that the N/V presentation always appeared as the twelfth test presentation.

The following sequence of events occurred during the test phase: .5-sec. warning buzzer —1-sec. exposure of a test card in the left window—9 sec. wait until the start of a new trial. Each *S* was told he would be shown stimulus cards singly in the left window which would be somewhat different from those shown previously. He was to name these, using one of the names he had used previously, whichever seem "most appropriate." If *S* did not give a name to the first presentation of a test stimulus he was shown the same stimulus card until he made a response. If two responses were given, the *S* was told to "give only one name please" and the name then given was recorded as the response. If an *S* spontaneously corrected a response this latter response was the one finally recorded. Before each test response the *S* was reminded to give the most appropriate name of the forthcoming presentation.

After responses had been given to all the test cards, *S*'s retention of each of the original name-class pairings was tested. The procedure was identical to the learning phase, except that only one example of each of the four concept classes was presented and no names were shown in the right window. The order of presentation of the stimulus types during the retention phase was randomized over all *S*s. The final retention trial concluded the experiment for an *S*.

RESULTS

Responses to single cues.—Of the responses to the single cues, 6.5% were inappropriate; i.e., were neither of the two responses associated with the cue during learning. This error rate is remarkably close to the 5.6% error rate obtained during the retention phase of the experiment.

Looking at the appropriate responses obtained when a 4 to 1 ratio of responses was predicted, we find that exactly 80% of the responses were the names of the high-frequency concept class, and when a 2 to 1 ratio was predicted, 68.2% of the responses were high-frequency names. Since each of these overall analyses includes four responses per *S*, for the purpose

TABLE 2

RESPONSES TO THE SINGLE CUES

Predicted Ratio	Response Stage	Obtained Frequency			χ^2 ($df = 1$) of Obtained from:		
		H	L	E	Matching	Chance Hypothesis	Other Ratio Hypothesis
4–1	First	117	29	14	.09	53.04***	11.96***
	Second	123	30	7	.03	56.33***	12.97**
	Third	117	35	8	.87	42.23***	7.30**
	Fourth	127	27	6	.59	64.94***	17.26***
	All	484	121	35			
2–1	First	106	36	18	4.32*	34.51***	2.54
	Second	97	50	13	.03	16.33***	18.04***
	Third	103	47	10	.27	20.91***	12.04***
	Fourth	98	55	7	.47	12.09***	24.72***
	All	404	188	48			

Note.—For this table and subsequent tables the "other" ratio of response is 2–1 when 4–1 is predicted and 4–1 when 2–1 is predicted.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

of a chi square analysis, one can look at each *S*'s first response when a 4 to 1 ratio was predicted, each *S*'s second response for this prediction and so on. This analysis of the effects of the relative frequency of reinforcement, presented in Table 2, provides firm support for the matching hypothesis.

Within the 4 to 1 relative ratio, the 8 to 2 and 4 to 1 absolute ratios of association per learning block produced 84.7% and 73.2% high-frequency responses, respectively. Similarly, within the 2 to 1 relative ratio, an absolute ratio of 8 to 4 produced 74.7% high-frequency responses compared to 61.4% high-frequency responses for the 2 to 1 absolute ratio. The results of separate analyses of each response stage for both predicted ratios were highly consistent with these tendencies to overshoot and undershoot matching as a function of the absolute frequency. However, in only three of the eight comparisons do the differences between absolute frequencies attain statistical significance ($p < .05$).

Responses to test presentations with too many cues.—In order to demonstrate the predictions derived from a component model note that in Table 1 the *S*s of Group G had the following frequencies of association during each learning block: CV-R_1, eight times; SV-R_2, four times; CN-R_3, twice; and SN-R_4, once. If upon presentation of Test Stimulus CNV the *S* randomly samples the elements previously associated with each of the three individual cues, we would predict 48.9% $R_1\left[\frac{80 + 0 + 66.7}{3}\right]$; 11.1% $R_2\left[\frac{0 + 0 + 33.3}{3}\right]$; 28.9% $R_3\left[\frac{20 + 66.7 + 0}{3}\right]$; and 11.1% $R_4\left[\frac{0 + 33.3 + 0}{3}\right]$.

A pattern model assumes that when *S*s are required to respond to either the SNV or CNV stimulus they should continue to sample the patterns of cues that occurred during learning (SN, SV, CN, or CV). While any

TABLE 3

FREQUENCIES OF THE VARIOUS NAMES GIVEN TO TEST STIMULI SNV AND CNV

Pattern Absolute Frequency	Absolute Learning Frequency				χ^2 ($df = 1$) of Obtained from Prediction of:			
	8	4	2	1	Component	Random Pattern	Weighted Pattern	Other Ratio
8–2	58	(3)	17	(2)	11.84***	22.41***	.33	3.84*
4–1	(9)	49	(9)	13	15.34***	20.90***	.04	4.30*
8–4	53	21	(5)	(1)	2.47	13.84***	.83	3.15
2–1	(7)	(6)	39	28	44.12***	1.81	2.18	19.93***

* $p < .05$.
*** $p < .001$.

response would be considered appropriate by the component model, only two of the four possible responses to any of the CNV or SNV stimuli would be considered appropriate by a pattern model. For example, when the *S*s of Group B respond to the CNV stimulus, the component model would predict that 53.4% of the obtained responses would be either an SN or an SV name, while the pattern model would predict none.

When the *S*s of Group B or D responded to a CNV stimulus, either CV or CN appeared eight times per learning block while the other appeared four times. For the sake of exposition we will say the *pattern* was associated in an 8 to 4 ratio for these *S*s. We can also speak of 2 to 1, 4 to 1, and 8 to 2 patterns. The tabulation of the various responses to the CNV and SNV stimuli, presented in Table 3, is in terms of these absolute frequencies. The entries in parenthesis are those responses that would be considered inappropriate by a pattern model. Since the results obtained with the CNV and SNV test stimuli were essentially similar they are combined in this analysis which provides little support for a component model.

If the cues are associated as patterns rather than individually, upon presentation of a stimulus having too many cues, it can be assumed that the patterns can be selectively sampled in either of two ways. Both types of pattern-sampling models predict few inappropriate responses but differ in their predictions of the ratio of appropriate responses. On the one hand the basis for pattern sampling can be randomly determined (random-sampling model). Thus, of the appropriate responses to the CNV test stimulus we would predict that 50% would be the names associated with CN, while 50% would be CV names.

On the other hand, we could predict that the probability of selective sampling either pattern is equal to the frequency of presentation of the pattern during learning (weighted-pattern model). Thus for the *S*s of Group G, responding to a CNV stimulus, we would predict that 80% of the appropriate responses would be CV names while 20% would be CN names.

Returning to Table 3, these results clearly support the weighted-pattern model.[3] This is shown by the low values of χ^2 in the "weighted pattern" column and the generally significant

---

[3] Following a suggestion by W. K. Estes a weighted-component model was derived in an attempt to account for the obtained response frequencies. However, this model was as fruitless as the previous (random) component model.

χ^2s in the "random pattern" column and the "other ratio" column.

Responses to presentations with both too many and too few cues.—It was assumed that a response to the NV stimulus involves two stages. The *S* first samples either the verb or the noun. Then, having sampled either, decides between the two appropriate responses associated with the word. Pitting the weighted-pattern model against the random-pattern model by testing the agreement between the relative frequencies of previous reinforcement and the frequencies of choice of N and V produced equivocal results.

Once the choice between the two parts of speech has been made, the matching hypothesis would predict that the relative frequencies of the various names to the part of speech would equal the relative frequencies with which the names were associated to the sampled part of speech during learning. It should be noted that in the analysis of these results, presented in Table 4, no response can be inappropriate. If, for example, *S* was, in fact, responding to the verb but gave a name associated with a noun, he would be included among those *S*s responding to the noun.

Despite the possibility of including errors as appropriate responses, where the matching model predicts a 4 to 1 ratio of names, 77.5% of the responses were high-frequency names and where the prediction was a 2 to 1 ratio of the responses were names, 68.8% of the responses were high-frequency names.

Responses to the intrinsically ambiguous cues.—These results demonstrated a tendency to undershoot matching when either a 4 to 1 or a 2 to 1 ratio of responses was predicted.

Response dependencies.—The data indicate that *S*'s responses to each cue singly, were independent of his responses to other types of cues. Similarly responses to SNV were independent of responses to CNV and vice versa. However, when *S*s gave the second response to a specific single cue there were significant dispositions to repeat more responses than would be expected under the assumption of independence.

TABLE 4

RESPONSES TO THE NV STIMULI AS A FUNCTION OF THE FREQUENCY OF LEARNING PRESENTATION

Predicted Ratio	Obtained Frequency		χ^2 ($df = 1$) of Obtained from:		
	H	L	Chance	Predicted	Other Ratio
4–1	62	18	24.20***	.31	4.25*
2–1	55	25	11.15***	.16	6.33*

* $p < .05$.
*** $p < .001$.

DISCUSSION

The success of the matching hypothesis in predicting the frequencies of responses to ambiguous cues is clearly demonstrated in the analysis of responses to the single cues. These results acquire further significance because of both the broad basis for class inclusion during learning and the fact that the single cues in the present experiment may have been presented as late as the eleventh test trial.

Although probability matching generally occurs, the absolute frequency of presentation also has an influence upon responses. Given any relative ratio of cue-name associations, there is a tendency for overshooting of matching with greater absolute frequencies and for undershooting of matching with smaller absolute frequencies. If it had been found that the frequency of responses for the cues whose absolute frequencies of presentation during each learning block were 2 to 1 and 4 to 1 undershot matching while the cues whose absolute frequencies were 8 to 4 and 8 to 2 exhibited matching, it would have been possible to argue that the obtained frequencies of responses to the two former cues were closer to chance because conditioning of the elements associated with these cues had not reached asymptote. However, since the frequencies of responses to the two latter cues overshot matching predictions, this line of reasoning seems inadequate unless the true situation is, as Edwards (1961) suggests, a tendency toward maximization.

The assumption that elements are associated with individual cues rather than patterns of cues receives little empirical support, since an assumption of patterned conditioning of elements leads to greater accuracy in predicting the percentage of inappropriate responses to presentations having too many cues.

The results with the SNV and CNV stimuli indicate that the frequencies of cue presentations affect the probabilities of selective sampling the elements associated with these cues. The *S*s do not sample patterns randomly after a history of differential reinforcement; instead the probability of sampling a pattern of cues tends to equal the frequency of occurrence of the pattern under conditions permitting unique class specification. This finding would support Straughan's (1953) suggestion that a "lag" may occur in the sampling process so that the probability of sampling a cue on one trial increases if it has been sampled on a previous trial.

Although the obtained frequencies of responses to the presentations having both too few and too many cues is not supportive of the weighted-sampling model, there was a clear tendency for cues appearing more frequently to be sampled more often. The inclusion of all responses as appropriate may have lead to the deviations from predictions. However, there is a more important reason why one cannot expect the weighted-sampling model to produce accurate predictions for the NV presentations. During learning *S*s were presented with patterns of cues and thus learned to sample the elements associated with patterns of cues on a differential basis. With the presentation of an NV test stimulus, an *S* is required to sample either of two single cues and consequently his sampling tendencies are to some extent independent of previous learning. The whole question of the effects of pattern presentations upon single cue sampling probability and vice versa requires further study. Despite these considerations, the results indicated that with this type of ambiguous stimulus, having both too few and too many cues for unique class specification, the matching hypothesis generated highly accurate predictions of response frequencies once sampling had taken place.

There are at least three reasons that could account for the obtained deviations from the matching prediction when *S*s responded to intrinsically ambiguous presentations. The first is that the tabulation of responses considered all responses to be appropriate. An analysis of the errors to the single cues indicated that the inclusion of all responses as appropriate would tend to make the obtained frequencies of responses nearer a chance distribution. Secondly, all the responses to the N/V stimuli were on the twelfth trial. In an analysis of the responses to the single cues there was a tendency for fewer high-frequency responses on a second trial with a particular cue type. Whether *S* responded to the N/V stimulus as a verb or a noun he was presented with a third trial with a particular part of speech, and we might expect a further decrease in the ratio of high-frequency responses with a resultant regression toward chance. Finally, there is the possibility that intrinsic ambiguity per se might lead to a deviation from matching predictions because of the simultaneous competition of four alternative responses.

The finding that successive responses to single cues of a specific type were not independent is in agreement with the findings of Binder and Feldman (1960), Estes (1960), and Edwards (1961). On the other hand it appears that an *S*'s response to an ambiguous presentation is independent of previous responses to differing ambiguous presentations; i.e., where associated responses differ. This finding confirms that of Schoeffler (1954) who found that ordinal position of test trials alone was not a significant variable in determining responses.

In brief, a probabilistic model of visual recognition and cue sampling appears to be very powerful when used to predict responses to ambiguous presentations.

REFERENCES

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.

BINDER, A., & FELDMAN, S. E. The effects of experimentally controlled experience upon recognition responses. *Psychol. Monogr.*, 1960, **74** (9, Whole No. 496).

EDWARDS, W. Probability learning in 1000 trials. *J. exp. Psychol.*, 1961, **62**, 385–394.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

SCHOEFFLER, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48**, 323–329.

STRAUGHAN, J. H. An application of statistical learning theory to an escape learning situation using human subjects. Unpublished doctoral dissertation, Indiana University, 1953.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Bureau of Publications, 1944.

(Received April 10, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 248–257

EFFECTS OF OMITTED VERSUS DELAYED UCS ON CLASSICAL EYELID CONDITIONING UNDER PARTIAL REINFORCEMENT[1]

JOHN W. MOORE AND ISIDORE GORMEZANO

Indiana University

The comparative effects in eyelid conditioning of 2 partial reinforcement techniques of UCS omission and UCS delay was investigated under 3 partial reinforcement schedules of 25, 50, and 75%. Under both procedures reinforced trials consisted of presentation of the UCS at a CS-UCS interval of 500 msec. It was found that: (a) UCS omission resulted in higher levels of conditioning and greater resistance to extinction than UCS delay; (b) shorter CR latencies were obtained under UCS omission than UCS delay; and (c) acquisition performance increased directly with percentage reinforcement, but reinforcement schedule failed to produce systematic effects on extinction. An associative and nonassociative interpretation was offered for the decremental effects of the UCS-delay procedure.

The purpose of the present investigation was to compare the acquisition and extinction of CRs in classical eyelid conditioning for 25, 50, and 75% partial reinforcement under the partial reinforcement techniques of UCS omission and UCS delay.

Eyelid conditioning studies involving partial reinforcement have employed the familiar technique of presenting the CS alone (i.e., without the UCS) on nonreinforced trials. A recent innovation involves presentation of the UCS on every trial, but on nonreinforced trials its onset is delayed for an interval of time shown by McAllister (1953) and Boneau (1958) to produce little or no conditioning. Because this UCS-delay technique permits presentation of the UCS on every trial, it should presumably eliminate the confounding, usually inherent in partial reinforcement designs, between percentage reinforcement and number of UCS occurrences. Therefore, recent studies of partial reinforcement in eyelid conditioning (Reynolds, 1958; Ross, 1959; Ross & Spence, 1960) have employed the UCS-delay technique in order to equate groups for motivational effects presumably related to UCS occurrences. However, the comparative effects of UCS omission and UCS delay have not received systematic assessment within the same experimental design. The need for such a direct comparison was emphasized by the results of Moore and Gormezano (1961) in which instrumental avoidance and classical eyelid conditioning were compared in a yoked-control procedure. In that study, a CR by an instrumental *S* on a given trial either avoided or delayed the UCS for himself and noncontingent partner. It was found that under the instrumental conditions, the UCS-omission contingency led to a higher level of responding than the UCS-delay contingency. In the two noncontingent groups (yoked to the

---

[1] This research was supported in part by a grant to the second author from the National Science Foundation, Grant G 16030. A brief report of the experiment was presented at the Midwestern Psychological Association, Chicago, May 1961.

two instrumental groups), the one receiving the delayed UCS responded at the higher level. This comparison was confounded, however, by the fact that this group, because of the yoking procedure, received the greater number of reinforced trials. The present investigation was designed, therefore, to compare the effects of the UCS-omission and UCS-delay techniques under several partial reinforcement schedules.

METHOD

Apparatus.—The apparatus has been described in detail elsewhere (Moore & Gormezano, 1961). The *S*'s room was sound treated and contained two identical enclosures which permitted 1 or 2 *S*s to be run at a time. The *S* sat on a straight back chair facing into a flat white, well-illuminated enclosure with his head comfortably positioned on a foam rubber chin rest. About 3 ft. in front of *S*'s eyes was a 10-cm. circular milk glass window. An abrupt change in hue (from pale blue to pale orange) and an increase of illumination (from .19 to .26 mL.) of this window served as the CS. The *S* wore an elastic headband supporting an air jet with a 2-mm. orifice directed at an upward angle and at a distance of $\frac{5}{8}$ in. from the right cornea. A 100-msec. puff of compressed nitrogen served as the UCS and was of sufficient intensity at the eye to support a 160-mm. column of mercury. The headband also supported a microtorque potentiometer coupled to *S*'s right upper eyelid by a light mechanical linkage. Movement of the lid produced a signal from the potentiometer which was amplified by a Brush dc amplifier and recorded by means of an ink writing Brush pen motor. A Western Union tape transmitter controlled the intertrial intervals and activated electronic timers controlling the duration of the CS and UCS, and the CS-UCS interval. The *E* and recording and stimulus programing equipment were housed in a sound treated room adjacent to the *S*'s room.

Subjects and design.—The *S*s were 48 men and 72 women from introductory psychology classes at Indiana University. Twenty *S*s were randomly assigned to each of the six cells of a 2 × 3 factorial design in which the two partial reinforcement techniques of UCS omission (O) and UCS delay (D), were orthogonal to three reinforcement schedules of 25, 50, and 75%. The random assignment of *S*s was restricted to requiring an equal ratio of men to women (8:12) in each of the cells. The two recording systems were also made orthogonal to the principal dimensions of the design in order to remove their contribution to the error variance.

The *S*s were randomly assigned to one of two independent random sequences of reinforced and nonreinforced trials within each reinforcement schedule. The randomization of reinforcement sequences was restricted by requiring that the first and last trials of acquisition be reinforced and the percentage of reinforcement in each four-trial block be equal to the percentage reinforcement schedule.

Procedure.—Instructions to *S*s were given over an intercommunication set and were the same as the "nonvoluntary" instructions stated fully in an earlier report (Gormezano & Moore, 1962). Each *S* received 80 acquisition and 20 extinction trials with randomly distributed intertrial intervals of 15, 20, and 25 sec. and mean of 20 sec. For *S*s run with the UCS-omission (O) procedure a nonreinforced trial consisted of presentation of the CS alone for 600 msec. Under the UCS-delay (D) procedure, a nonreinforced trial consisted of presentation of the CS for 2600 msec. terminating together with the UCS. Under both procedures reinforced trials were defined as a 600-msec. CS terminating together with the UCS (i.e., a .5-sec. CS-UCS interval). In extinction all trials were nonreinforced trials of the UCS-omission type.

Response measures.—In acquisition any deflection from the baseline of at least 1 mm. occurring in the 150–550 msec. interval after CS onset was scored as a CR. This interval was extended to 600 msec. in extinction.

RESULTS

Figure 1 presents performance curves for the six experimental groups in acquisition and extinction. The initial points on all acquisition curves are the mean percentage CRs on the first trial and the remaining points are the mean percentage CRs on successive blocks of 10 trials. The extinction curves are plotted in 5-trial blocks. In acquisition, Groups 75-O and 50-O attained asymptotes of approximately 85% within 40 trials.

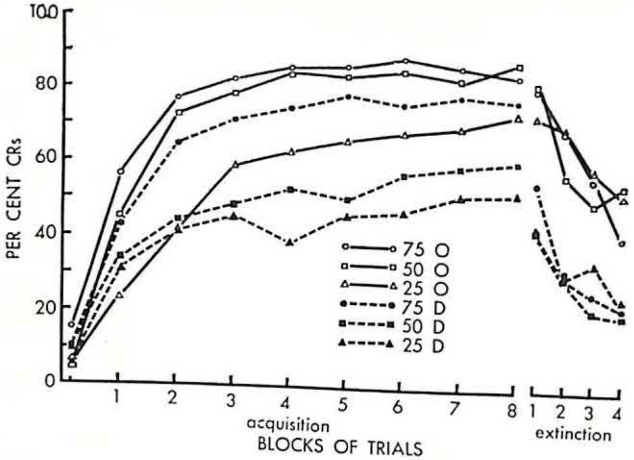


FIG. 1. The percentage of CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

The curves for the remaining groups rose gradually and do not appear to have stabilized within the 80 conditioning trials. The mean percentage CRs for Groups 25-O, 50-O, and 75-O in acquisition were 56.6, 78.2, and 83.4%, respectively, whereas the means were 42.4, 49.7, and 71.2% for Groups 25-D, 50-D, and 75-D, respectively. Therefore, within all values of the partial reinforcement schedule, the O groups responded at higher levels than did the corresponding D groups, and within reinforcement technique, level of responding increased with increasing percentage reinforcement.

Extinction curves are presented on the right-hand side of Fig. 1. It is clear that the O groups responded at considerably higher levels in extinction than did the D groups. Moreover, within the first five extinction trials, the level of responding for the D groups dropped approximately 20% from the last block of acquisition trials whereas there was little or no immediate decrement among the O groups. Within the UCS-omission and UCS-delay procedures, reinforcement schedule apparently did not differentiate among the groups in extinction in any simple way. The mean percentage CRs in extinction for Groups 75-O, 50-O, and 25-O were 64.3, 62.5, and 64.5%, respectively, and for Groups 75-D, 50-D, and 25-D the means were 30.5, 24.6, and 28.4%, respectively. Thus, the two partial reinforcement techniques were widely separated in extinction while reinforcement schedule exhibited little systematic effect.

The frequency distribution of response latencies in acquisition is shown in Fig. 2 for all groups. The O groups are depicted on the left and the D groups on the right with reinforcement schedule as the parameter. The responses distributed about a mode at 75–100 msec. are reflex responses to the CS (Grant, 1943), and it is of particular interest to note that more of these alpha responses were obtained under 25 and 50% reinforcement than under the 75% reinforcement schedule. The low frequency of responses in the 125–200 msec. interval for all groups indicates that precautions to maintain light adaptation were successful in eliminating beta responses (Grant &

Norris, 1946). Examination of the principal modes indicates that partial reinforcement technique had a considerable effect on CR latency. The D groups showed a flatter distribution than the O groups with a comparatively large proportion of responses at the long latency end of the distribution. Inspection of the figure further indicates that reinforcement schedule had little systematic effect on CR latency. The mean CR latency in acquisition was computed for each *S* and averaged over *S*s to yield mean CR latencies for Groups 75-O, 50-O, and 25-O of 335, 337, and 346 msec., respectively, and for Groups 75-D, 50-D, and 25-D of 380, 391, and 381 msec., respectively.

The frequency distribution of response latency in extinction is shown in Fig. 3. The first, third, and fifth distributions from the top of the figure are the UCS-omission (O) groups; the second, fourth, and sixth distributions are the UCS-delay (D) groups with percentage reinforcement in acquisition as the parameter. It is clear from the figure that the O groups gave more CRs in extinction than did the D groups. It is also apparent that the modes of the respective distributions did not change to any great extent from acquisition to extinction. Mean CR latencies in extinction were obtained, as in acquisition, by averaging the means of individual *S*s. These means were 331, 336, and 335 msec. for Groups 75-O, 50-O, and 25-O, respectively, and 383, 391, and 391 msec. for Groups 75-D, 50-D, and 25-D, respectively.

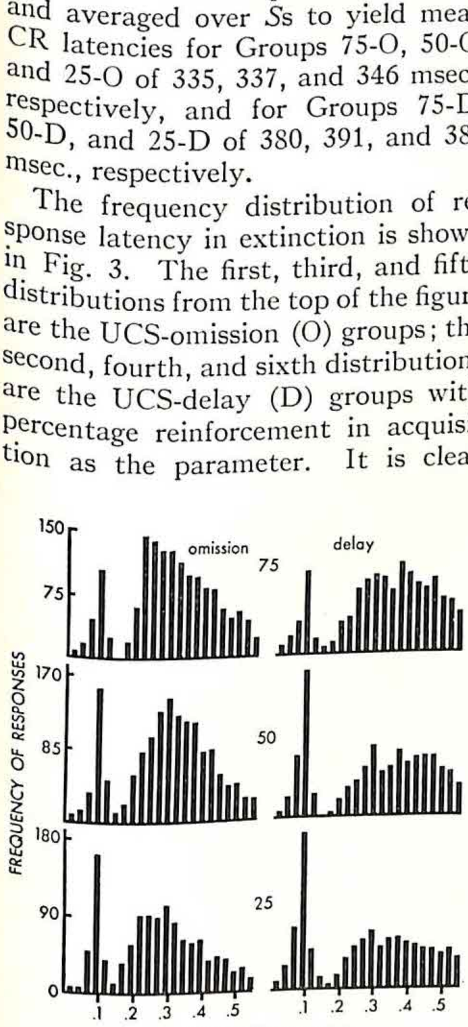


FIG. 2. Latency distributions of all eyelid responses in acquisition.

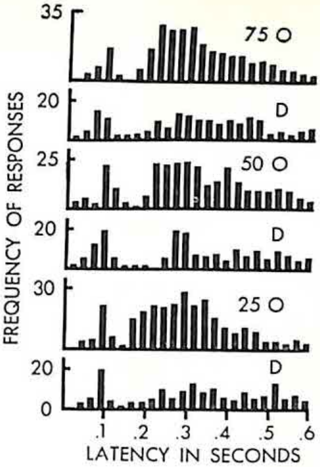


FIG. 3. Latency distributions of all eyelid responses in extinction.

The arc-sine transforms of individual percentage CRs and the individual mean CR latencies in acquisition and extinction were subjected to analysis of variance. Table 1 presents the summary of these analyses for both the percentage and latency measures. For each response measure the acquisition analysis appears in the left-hand column and the extinction analysis in the right-hand column. Except for the error *MS*s which appear in parentheses at the bottom of each column, only the *F* ratios are presented. In Row 1, the source of variation tested is the effect of partial

TABLE 1

ANALYSES OF VARIANCE OF ARC-SINE PERCENTAGE CRs AND MEAN CR LATENCY IN ACQUISITION AND EXTINCTION

Source of Variation	*df*	*F* Ratios			
		Percentage		Latency	
		Acquisition	Extinction	Acquisition	Extinction
Partial reinforcement technique (T)	1	12.31***	27.63***	17.16***	9.91***
Reinforcement schedule (S)	2	9.35***	.08	.15	.82
Recording system (R)	1	.00	.41	.57	.59
T × S	2	.86	.07	.28	.51
T × R	1	1.86	.04	.36	.04
S × R	2	2.98	.51	1.86	.44
T × S × R	2	2.83	1.60	1.16	.26
Error *MS*	108	(310.63)	(501.59)	(136.89)	(237.01)

*** $p < .001$.

reinforcement technique. According to the F values, there were significant differences in the percentage and latency of CRs in both acquisition and extinction.

Binary comparisons (two-tailed t tests based on the error MS of the analysis of variance) of the O with the D groups for each reinforcement schedule indicated that the superior performance of Group 50-O over Group 50-D ($p < .01$) was the primary contributor to the F value obtained for the effects of reinforcement technique upon acquisition percentage scores. Comparisons of 25-O with 25-D and 75-O and 75-D failed to reveal statistically reliable differences. The value of F obtained for the effects of reinforcement technique on extinction percentage scores was due to the greater resistance to extinction of the O over the D groups for each reinforcement schedule (all p's $< .01$). Furthermore, differences between O and D groups in extinction, although greatly attenuated, remained reliable after the linear regression of extinction scores on acquisition scores had been taken into account by an analysis of covariance ($F = 12.98$, $p < .001$). Binary comparisons indicated that the largest contributions to the F value obtained for the effects of reinforcement technique on acquisition latency scores came from the reliably shorter latencies of the O groups to the D groups under 50 and 75% reinforcement (viz., 50-O vs. 50-D, $p < .01$; 75-O vs. 75-D, $p < .05$). For extinction, t tests indicated that the principal contributions to the F observed for the latency measure came from the reliably shorter CR latencies of the O groups to the D groups under 25 and 75% reinforcement (p's $< .05$).

As is shown in Row 2 of Table 1, reliable differences due to reinforcement schedule were obtained in the analysis of percentage CRs in acquisition but not in the analysis of percentage CRs in extinction or in the analyses of CR latency, in acquisition and extinction. In addition, when the linear regression of extinction percentage scores on acquisition percentage scores was taken into account by an analysis of covariance, differences due to reinforcement schedule failed to reach statistical significance ($F = 1.68$). The binary comparisons

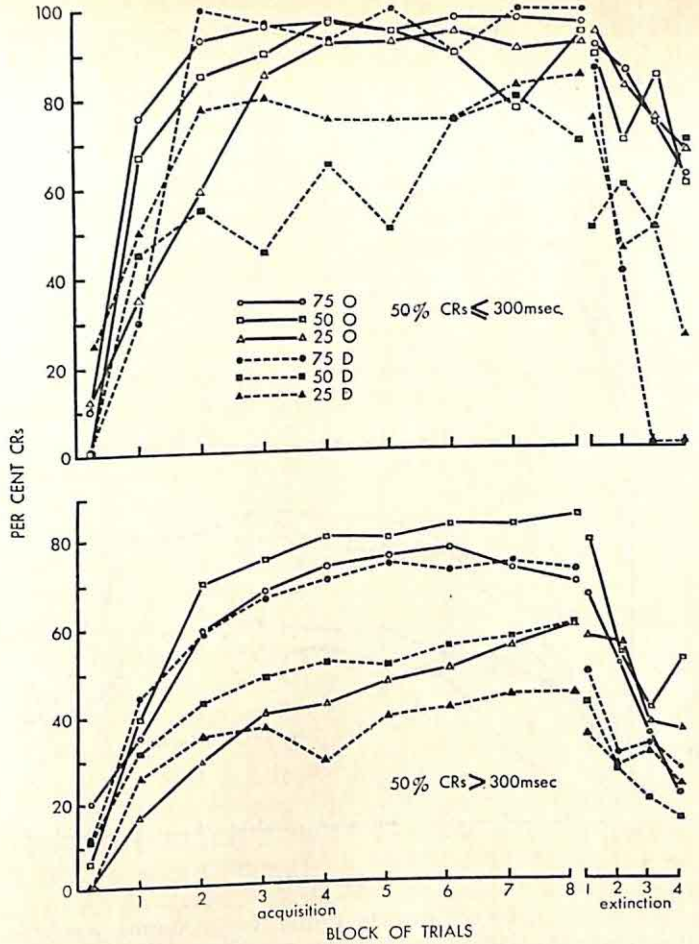


FIG. 4. The percentage of CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction for *S*s classified as V (upper graph) and NV (lower graph) with a CR latency criterion of 300 msec.

among reinforcement schedules (i.e., with O and D groups pooled for the comparisons) indicated that the largest contributions to the *F* value obtained for reinforcement schedule in acquisition came from the reliably higher level of responding of the 75% reinforcement group over the 25% ($p < .01$) and 50% ($p < .01$) groups.

Row 3 of Table 1 tests for variation due to the recording systems and Rows 4 through 7 test for the possible interaction of main effects. The *F* values indicate that the sources of variation in Rows 3 through 7 failed to reach significance for either the percentage CRs or CR latency measures in either acquisition or extinction.

An attempt was made to separately analyze the data of so-called "voluntary" (V) and "nonvoluntary" (NV) responders.[2] Those *S*s who gave 50%

---

[2] An analysis for "voluntary" responders was considered necessary by the consultant. For statistical analyses of data so classified, the arc-sine transform of percentage CRs was employed: All t tests were two-tailed and only those significant at $p \leqslant .05$ are reported.

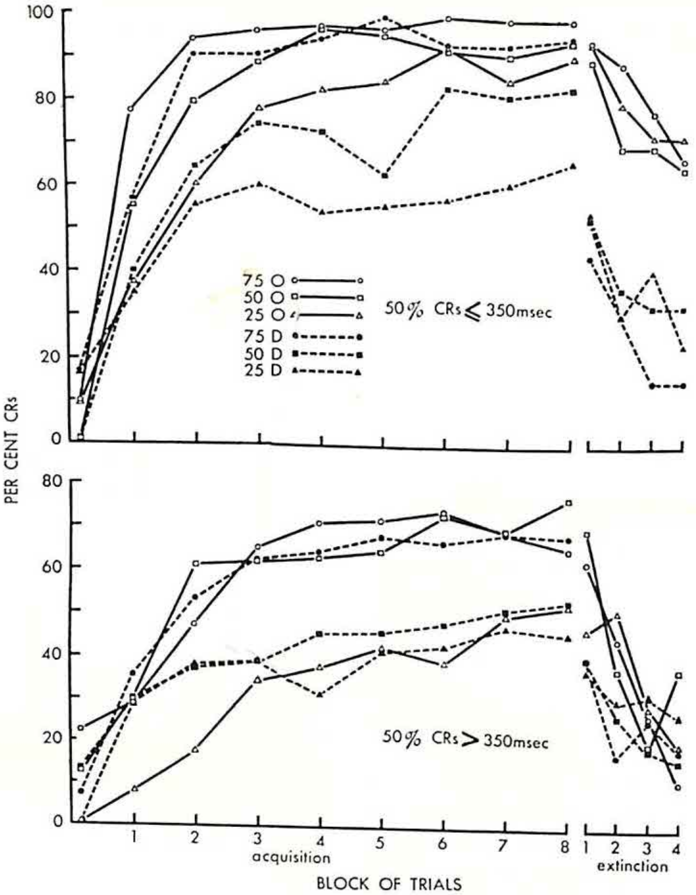


FIG. 5. The percentage of CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction for *S*s classified as V (upper graph) and NV (lower graph) with a CR latency criterion of 350 msec.

or more of their CRs with a latency of 300 msec. or less were designated as V *S*s (Spence & Ross, 1959). The 31 *S*s so classified were distributed among the six experimental groups as follows: 25-D, $N = 4$; 50-D, $N = 2$; 75-D, $N = 3$; 25-O, $N = 8$; 50-O, $N = 4$; and 75-O, $N = 10$. The acquisition and extinction curves for V and NV *S*s are presented in the upper and lower graphs, respectively, of Fig. 4. With the exception of the 75-O vs. 75-D comparison, V *S*s reached higher levels of responding in acquisition under the O condition and pooling over reinforcement schedules, demonstrated a significantly greater resistance to extinction than under the D condition, $t(29) = 4.83$, $p < .001$. The acquisition and extinction curves for the 89 NV *S*s (lower graph) closely resemble those presented in Fig. 1. The main differences are an overall lower level of responding and the occurrence of a position reversal between Groups 75-O and 50-O in acquisition. Nevertheless, an O vs. D comparison revealed a significantly higher level of responding under the O condition in both acquisition, $t(87) = 1.96$, $p < .05$ and extinction, $t(87) = 2.77$, $p < .01$; and in particular, Group 50-O was significantly superior to Group 50-D in acquisition, $t(32) = 2.88$, $p < .01$. Comparisons among reinforcement schedules indicated significant differences for the 75% vs. 25%, $t(52) = 3.94$, $p < .001$ and 50% vs. 25% comparisons, $t(60) = 3.00$, $p < .01$. Furthermore, pooling over reinforcement schedule, the observed higher level of responding of V over NV *S*s (i.e. upper vs. lower graph) was significant for the following comparisons: V-O vs. NV-O in both acquisition, $t(58) = 4.39$, $p < .001$ and extinction, $t(58) = 3.70$, $p < .01$; and V-D vs. NV-D in acquisition, $t(58) = 2.64$, $p < .02$.

Since a ready signal was not employed in the present investigation, a 300-msec. latency criterion would presumably be a conservative criterion for recovering so-called voluntary responders (Hartman, Grant, & Ross, 1960). Accordingly, Fig. 5 presents separate acquisition and extinction curves for V and NV *S*s defined by using a 350-msec. criterion (i.e. *S*s who gave 50% or more of their CRs with a latency of 350 msec. or less were designated as V *S*s). The 51 V *S*s so classified were distributed among the six experimental groups as follows: 25-D, $N = 6$; 50-D, $N = 5$; 75-D, $N = 6$; 25-O, $N = 11$; 50-O, $N = 12$; and 75-O, $N = 11$. Analysis of performance of V *S*s, presented in the upper graph of Fig. 5, indicated a significantly higher level of responding for the O over D reinforcement technique in both acquisition, $t(49) = 2.11$, $p < .05$ and extinction, $t(49) = 4.01$, $p < .001$. Summing over partial reinforcement technique revealed that the 75% reinforcement schedule was significantly superior to the 50%, $t(32) = 3.14$, $p < .01$ and 25% reinforcement schedule, $t(32) = 3.48$, $p < .01$ in acquisition. The data for NV *S*s, presented in the lower graph of Fig. 5, reveals that most order relationships were retained among groups, but an O vs. D comparison failed to reveal significant differences in either acquisition or extinction. For partial reinforcement schedule in acquisition, however, 25% reinforcement was significantly inferior to 75%, $t(44) = 3.23$, $p < .001$ and 50% partial reinforcement, $t(44) = 2.28$, $p < .05$. In V vs. NV comparisons (i.e. upper vs. lower graph) significant differences were obtained for V-D vs. NV-D in both acquisition, $t(58) = 2.67$, $p < .01$ and extinction, $t(58) = 2.32$, $p < .05$; and V-O vs. NV-O in both acquisition, $t(58) = 6.14$, $p < .001$ and extinction, $t(58) = 5.93$, $p < .001$.

DISCUSSION

The major findings of this investigation were: (*a*) partial reinforcement under the UCS-omission procedure resulted in higher levels of conditioning and greater resistance to extinction than were obtained with UCS delay; (*b*) level of responding in acquisition increased directly with percentage reinforcement, but reinforcement schedule produced little or no differences in extinction; and (*c*) shorter CR latencies were obtained with UCS omission than with UCS delay.

The finding that a UCS-delay procedure produced a performance decrement compared with UCS omission corresponds with the observations of Moore and Gormezano (1961) for instrumental conditioning and further suggests that the direction of differences between yoked-control groups observed in that study (i.e., UCS delay produced a higher level of responding than UCS omission) was due to the higher reinforcement schedule and not to partial reinforcement technique. In addition, it is clear from previous uses of the UCS-delay procedure, that it has been assumed that a delayed UCS has the same effect on habit strength as ordinary nonreinforcement. Specifically, the UCS-delay procedure has been employed

TABLE 2

Conditional Proportions for UCS-Delay Groups of a Response and No Response in the Interval of Delay Given a CR or No CR in the Criterion Interval

Experimental Condition	Response		No Response	
	CR	No CR	CR	No CR
75-D	.410	.138	.361	.091
50-D	.344	.225	.156	.275
25-D	.254	.322	.171	.253

in studies interested in comparing continuous (100%) reinforcement with partial (50%) reinforcement (Reynolds, 1958; Ross, 1959; Ross & Spence, 1960). These studies have, in fact, demonstrated the usually observed superiority of continuous over 50% partial reinforcement in eyelid conditioning. The present results would seem to indicate, however, that decremental effects attributable to nonreinforcement alone can be spuriously overestimated by using a UCS-delay procedure.

The performance decrement in acquisition obtained with the delayed UCS can perhaps be accounted for, in terms of an associative factor of concurrent conditioning in the delayed interval of a response which competes with the CR in the shorter CS-UCS interval. Some independent support for this interpretation is provided by the relatively long CR latencies obtained with the delayed UCS. Conditioning in the interval of delay may tend to displace criterion CRs toward the delayed UCS. Somewhat stronger support for a competing response interpretation is provided by the obtained conditional proportions of a response in the delay interval given that a CR did, or did not occur in the criterion interval. The proportions for the UCS-delay groups are presented in Table 2. Of particular relevance is the observation that an increasing proportion of UCS-delay trials (i.e., decreasing percentage reinforcement) produced an increasing proportion of trials on which a response occurred in the delay interval but not in the criterion interval. A difficulty with this interpretation, however, was a failure to find an increase over conditioning trials, in the percentage of responses in the delay interval. It is possible that some nonassociative factor is responsible for the decrement. If for example, emotionally based drive r_e (Spence, 1958) decreases as a function of number of UCS occurrences due to adaptation of r_e, the relevant portions of the data could be accounted for.

Analysis of covariance indicated that the greater resistance to extinction obtained with UCS omission cannot be completely accounted for in terms of a continuation in extinction of differences accruing in acquisition. The comparatively rapid rate of extinction for the UCS-delay groups in the present experiment was probably due to the procedural change of shifting from UCS delay in acquisition to straight UCS omission in extinction. It would appear that any of the current hypotheses which explain resistance to extinction in terms of stimulus change can account for the obtained results (Lewis, 1960).

The finding that level of responding in acquisition increased monotonically with percentage reinforcement is a conventional result. Furthermore, the effects of partial reinforcement schedule on the total number of CRs in extinction seems to correspond with the findings of Hartman and Grant (1960) and Froseth and Grant (1961).

The observation that CR latency varied with partial reinforcement technique has methodological implications for eyelid conditioning studies which attempt to identify so-called voluntary responders on the basis of response latency (Spence & Ross, 1959). When a 300-msec. criterion was applied to the present data, reliable superiority of UCS omission over the UCS-delay procedure was obtained for NV but not V *S*s. On the other hand, a 350-msec. criterion indicated the complementary result, viz., reliably higher responding under UCS omission than UCS delay for V *S*s but not for NV *S*s. Since 22 of the 31 *S*s classified as voluntary responders with the 300-msec. criterion and 34 of the 51 *S*s so classified with the 350-msec. criterion received the UCS-omission procedure, the classification of *S*s was not independent of reinforcement technique. Under the assumption that *S*s are a priori voluntary responders one would expect an equal proportion of these *S*s under each reinforcement technique. Under each criterion this expectation must be rejected (for 300 msec., $\chi^2 = 5.32$, $p < .025$; for 350 msec., $\chi^2 = 6.42$, $p < .025$). It could be assumed, of course, that UCS omission simply generated more voluntary responders than the UCS-delay procedure. However, in the absence of empirical evidence that the response latency criterion employed in the classification of *S*s as voluntary responders is *not* a function of the independent variables employed, it is not implausible that *S*s are being classified for demonstrating behavior functionally relating the conditioning process to reinforcement technique. Specifically, the alternative interpretation exists that UCS omission produced better conditioning and that *S*s classified as voluntary responders on the basis of response latency are simply those with higher learning rate parameters. Evidence for such an interpretation has been provided in a previous study (Gormezano & Moore, 1962) and is supported in the present investigation by the observation that V *S*s on the whole, demonstrated higher levels of responding than NV *S*s in both acquisition and extinction.

REFERENCES

BONEAU, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

FROSETH, J. Z., & GRANT, D. A. Influence of intermittent reinforcement upon acquisition, extinction, and spontaneous recovery in eyelid conditioning with fixed acquisition trials. *J. gen. Psychol.*, 1961, **64**, 225–232.

GORMEZANO, I., & MOORE, J. W. Effects of instructional set and UCS intensity on the latency, percentage, and form of the eyelid response. *J. exp. Psychol.*, 1962, **63**, 487–494.

GRANT, D. A. Sensitization and association in eyelid conditioning. *J. exp. Psychol.*, 1943, **32**, 201–212.

GRANT, D. A., & NORRIS, E. B. Dark adaptation as a factor in the sensitization of the Beta response of the eyelid to light. *J. exp. Psychol.*, 1946, **36**, 390–397.

HARTMAN, T. F., & GRANT, D. A. Effect of intermittent reinforcement on acquisition, extinction, and spontaneous recovery of the conditioned eyelid response. *J. exp. Psychol.*, 1960, **60**, 89–96.

HARTMAN, T. F., GRANT, D. A., & ROSS, L. E. An investigation of the latency of "instructed voluntary" eyelid responses. *Psychol. Rev.*, 1960, **7**, 305–311.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

MCALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 416–422.

MOORE, J. W., & GORMEZANO, I. Yoked comparisons of instrumental and classical eyelid conditioning. *J. exp. Psychol.*, 1961, **62**, 552–559.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

ROSS, L. E., & SPENCE, K. W. Eyelid conditioning performance under partial reinforcement as a function of UCS intensity. *J. exp. Psychol.*, 1960, **59**, 379–382.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received April 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 258–264

A GUESSING STRATEGY WITH THE ANTICIPATION TECHNIQUE[1]

LLOYD R. PETERSON, CHARLES L. BREWER,[2] AND RICHARD BERTUCCO[3]

Indiana University

The capacity of Ss to utilize memory for recently presented responses to maximize guessing success was studied in 2 experiments. A serial anticipation technique was used with a list of 9 words whose order changed on each trial in Experiment I. In Experiment II the same words were used in a paired-associate task. Ss tended to give responses which had not previously been presented on the current trial. Correct responses increased significantly toward the end of a trial. Omissions increased from beginning to end of a trial. It was considered that a guessing strategy could produce the latter portion of the serial error curve, and in any list-learning inflate estimates of associative strength obtained by scoring anticipations.

There is a growing body of evidence from short-term retention studies that probability of a correct recall varies as a function of the recency of the prior presentation (Peterson, Peterson, & Miller, 1961; Peterson, Saltzman, Hillner, & Land, 1962). Similarly, recognition experiments have found that the accuracy with which an item can be identified as having occurred previously in the experiment is directly related to recency (Shepard & Teghtsoonian, 1961). The relationship seems to be reversed in the case of the rote learning of lists, where the most recently presented response is not the response most likely to occur as an erroneous anticipation. In serial learning errors have been analyzed as remote associations, and forward associations have been found to be stronger than backward associations (Bugelski, 1950). The present experiments investigate the possibility that the occurrence of remote associations is strongly influenced by a guessing strategy on the part of the *S*s. In the ideal case, after the *S* had learned what the responses were, if he could remember what items had already appeared on a given trial, then he could confine his guesses to the remaining items. Instead of the proportion correct due to guessing being equal for each position in the list, the guessing probability would be equal to the reciprocal of the number of items which had not yet appeared on that trial. Thus, for the last position there would be no errors, for the next last position the guessing probability would be one-half, and so on. The number of errors would be a monotonic decreasing function from beginning to end of the trial. It is an empirical question to what extent the *S* could approximate this ideal state of affairs.

Two experiments were designed to provide tests of the capacity of *S*s to utilize memory of responses previously presented on a given trial to maximize the probability of a correct anticipation. The initial experiment presented a list of nine familiar words over a series of trials. The anticipation technique was used after the manner of serial rote learning. However, the order of words was different on each trial, so that learning in the traditional sense could not occur. In a second experiment the same words were used as responses in a paired-associate learning task.

---

[1] This research was facilitated by National Science Foundation Grant G12917, for which the senior author is principal investigator.

[2] Now at University of Arkansas.

[3] Now at Wagner College.

EXPERIMENT I

Method

Subjects.—The *S*s were 30 undergraduates enrolled in an introductory psychology course at Indiana University for whom participation in experiments was a course requirement. Data from 1 additional *S* were discarded because more responses were made than the *E* could record.

Apparatus and materials.—Presentation was by a memory drum mounted in a 3 × 4 ft. black screen. A sliding shutter permitted exposure of nine columns of items without change of tape. Nine monosyllabic stimulus words from the Minnesota Word Association norms (Russell & Jenkins, 1954) were used, each having the characteristic that it did not occur as a response to any of the other stimulus words more often than three times in 1008 *S*s.

Procedure.—The words were arranged in 18 orders such that each word appeared equally often in all positions and preceded and followed every other word an equal number of times. Words were presented at a 4-sec. rate, and 16 sec. separated the end of one trial from the asterisks signaling the beginning of the next trial.

The following instructions were read to each *S*:

> In a moment there will be an opening in the screen and a list of words will be shown to you, one at a time. Just watch the words as we go through the list once, so that you will know what the words are like. Ready?

After a single presentation of the list, the instructions continued:

> Notice that a word occurs only once in the list. That will always be true. The order of the words will change, but a word doesn't repeat until the whole list is shown again. Now, this next time through the list try to guess ahead of time what words are coming next. A row of stars is the signal that the first word is coming up. When you see the stars, guess what the first word will be. While you are looking at the first word, guess what the second word will be, and so on through the list. You can expect to make many mistakes, but keep on trying. Here we go, guess.

After the second presentation of the list, further instructions were read which were intended to produce multiple responses in each anticipation interval. It was considered possible that *S*'s first and second responses might be differently related to recency of previous presentation. The instructions were:

> That's fine. Now, we're interested in how people think in this situation, so we'd like to have you think out loud. As soon as the drum stops at a word, say the first word you think of, outside of the one you're looking at, of course. Then, when you think of a better guess, say that word next. Make as many guesses in each position as you can. Do you have any questions? We'll go through the list a number of times, and please save any questions until the end. Now we'll begin.

The experiment then proceeded through the remaining 16 trials without interruption.

Results and Discussion

The instructions to give multiple responses proved to be ineffective; few *S*s gave more than one response within a single 4-sec. anticipatory period. In cases where multiple responses occurred only the last response was scored. In the discussion that follows, Anticipation n (A-n) will refer to *S*'s response as he looked at the n-first word in the list. Anticipation 1 (A-1) will refer to *S*'s response as he looked at the asterisks preceding the first word. Position n (P-n) will refer to the source of a response that *S* made. Thus, if on any anticipation *S* responded with a word which was fourth in the list on that trial, the source of the response will be called Position 4 (P-4).

The distributions by position of *S*'s voiced responses over Trials 3–18 are shown in Fig. 1. There was a general tendency for probability of emission of a word to be directly related to the

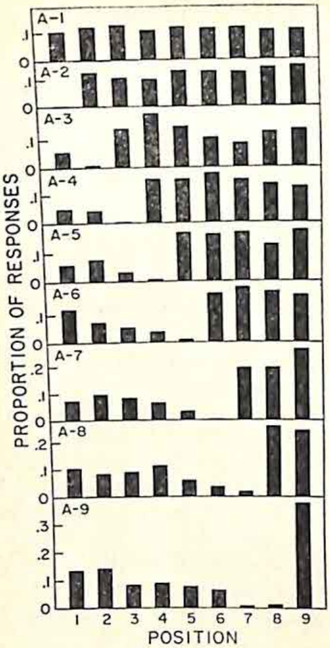


FIG. 1. Distributions of responses by position.

length of time since its last appearance in the drum aperture. The words emitted with highest frequency were those which had not yet appeared on that trial, and *S*s divided their guesses randomly among them.

The tendency to give responses that had not yet been presented during the current trial is analyzed in Table 1. Responses at each anticipation were scored as *Forward* anticipations if in the *E*'s list on that trial they occurred after the anticipation. They were called *Backward* if they had been presented prior to the anticipation on that trial. Responses read from the drum were not included in this analysis. Scores for each *S* at each anticipation were obtained by summing over Trials 3–18 and dividing by the number of different positions represented in the summation. Thus, at A-7, for instance, the divisor was 3 for the Forward category and 5 for the Backward category. Group means of these scores are listed in Table 1. Successive means in the Forward category increase from A-1 to A-9 with no reversals. Applying the binomial theorem with the assumption that an increase would be equally as likely as a decrease, the

TABLE 1

MEAN NUMBER OF FORWARD AND BACKWARD RESPONSES WEIGHTED FOR OPPORTUNITY OF OCCURRENCE

Exp.	Position	Anticipation								
		1	2	3	4	5	6	7	8	9
I	Forward $\bar{x}$ $SE_{\bar{x}}$	 1.67 .06	 1.89 .07	 2.01 .04	 2.26 .05	 2.43 .09	 2.49 .12	 2.79 .19	 2.93 .24	 3.77 .39
	Backward $\bar{x}$ $SE_{\bar{x}}$			 .83 .18	 .72 .12	 .84 .12	 .99 .09	 .90 .07	 .95 .09	 .83 .09
II	Forward $\bar{x}$ $SE_{\bar{x}}$	 1.09 .09	 1.16 .10	 1.31 .09	 1.45 .10	 1.40 .14	 1.34 .19	 1.42 .19	 1.87 .30	
	Backward $\bar{x}$ $SE_{\bar{x}}$		 .06 .04	 .21 .02	 .30 .06	 .42 .05	 .60 .08	 .66 .07	 .64 .08	 .69 .09

TABLE 2

MEAN PROPORTIONS OF CORRECT RESPONSES AND OMISSIONS

Exp.	Measure	Trials	Anticipation								
			1	2	3	4	5	6	7	8	9
I	Correct	3–10	.10	.12	.10	.13	.17	.12	.10	.17	.18
		11–18	09	.11	.17	.15	.14	.17	.21	.21	.29
	Correct After runs	3–10		.01	.01	.03	.03	.03	.04	.05	.06
		11–18		.00	.02	.03	.00	.04	.06	.07	.13
	Omissions	3–10	.04	.04	.06	.10	.13	.18	.31	.33	.46
		11–18	.02	.01	.03	.02	.02	.06	.08	.19	.32
II	Correct	2–9	.19	.23	.25	.20	.24	.22	.21	.21	.24
		10–18	.36	.40	.40	.40	.40	.44	.41	.42	.43
	Correct After runs	2–9		.04	.04	.02	.05	.04	.04	.05	.05
		10–18		.04	.01	.04	.03	.03	.05	.03	.06
	Omissions	2–9	.23	.21	.21	.25	.31	.31	.35	.38	.38
		10–18	.18	.19	.15	.18	.19	.20	.22	.23	.29

probability of eight successive increases is $(\frac{1}{2})^8$, or less than .004. A comparison of Forward to Backward means in A-3 to A-9 shows that Forward responses predominate at each anticipation. Two-tailed binomial sign tests at each of these anticipations where both Backward and Forward anticipations could be obtained were significant at the .01 level in every case. Means for the Backward anticipations do not appear to differ from one another save by chance.

There resulted a positive acceleration in the number of correct anticipations at the later positions of a trial, as can be seen in Table 2. Assuming that each response was equally likely, .11 of the anticipations would be correct by chance. However, chance was exceeded in the later anticipations of a trial. Over Trials 3–10, 21 of 30 *S*s were above .11 on A-4 and A-8 ($p < .05$). On A-5 and A-9 23 or more *S*s were above .11 ($p < .01$). Over Trials 11–18, 23 or more *S*s exceeded .11 on A-3, A-4, A-6, A-7, A-8, and A-9 ($p < .01$), and on A-5 22 *S*s were above .11 ($p < .05$). There was a tendency for *S*s to improve over the session, as evidenced by the fact that 24 *S*s made more correct anticipations during Trials 11–18 than during Trials 3–10 ($p < .01$).

Examination of omissions (Table 2) indicates that *S*s made fewer responses as a trial proceeded from A-1 to A-9. During Trials 3–10 28 *S*s made fewer responses on A-9 than on A-1 ($p < .01$). During Trials 11–18 23 *S*s made fewer responses on A-9 than on A-1 ($p < .01$). In contrast, as the *session* progressed, *S*s made more responses. Summing over all anticipations, 25 *S*s made more responses in Trials 11–18 than in Trials 3–10 ($p < .01$). It seems reasonable to conclude that *S*s recognized that certain responses were erroneous because of recent occurrence, and that they refrained from speaking when a better guess did not occur to them.

A decrease in total responding as a trial progresses has previously been noted in serial rote learning by Deese and

Kresse (1952). However, in their paced experiment after reaching a peak omissions decreased toward the end of the list. Any of several differences in the two experiments may be the basis of the discrepancy. A 2-sec. anticipation period was used in the Deese and Kresse experiment, and associative learning occurred. In their unpaced conditions omissions reached a plateau which extended to the end of the trial.

Further examination of the data of the present experiment indicated that sometimes *S*s repeated guesses on successive anticipations. However, the *S*s did not seem to have adopted the strategy of repeating the same guess until it proved correct. Seventy-seven percent of all runs were only two anticipations long. If only the runs ending in failure are examined, 85% were two anticipations in length. Table 2 lists the number of correct anticipations that marked the end of a run of two or more identical responses. These runs would seem to account for part, though not all, of the increase in correct anticipations from beginning to end of a trial.

The systematic trends exhibited in this experiment suggest that the latter portion of the serial-position error curve may be at least in part the result of a guessing strategy. A rough estimate of the adequacy of a guessing strategy to produce this result can be obtained by noting that in this experiment approximately half as many errors occurred in the final position as in the positions just past the center of the list. Several serial list learning experiments show comparable ratios (Bugelski, 1950; Deese & Kresse, 1952).

It was concluded that in any list learning where anticipation is followed by correction, and where the responses are used but once in a trial, avoidance of recently presented responses should increase the number of correct anticipations. Since the use of such a strategy would be reinforced, it would be expected that it would be used more frequently as a session progressed. Further investigation of the strategy was undertaken in a paired-associate learning experiment.

EXPERIMENT II

Method

The same nine words were again used as responses, but they were now paired with stimuli consisting of three-digit numbers. The numbers were of high similarity, since only the digits 6–9 were used. Responses appeared in the same orders as in the previous experiment over 18 trials. The stimulus member of a pair appeared in the drum for 2 sec. followed by the response (alone) for 2 sec. Sixteen seconds elapsed between trials. Thirty-one introductory students served as *S*s. The following instructions were read to each *S* initially:

> When we begin, a list of numbers and words will appear one by one in this small window. First you will see a number and then a word. Say each word aloud as soon as you see it and try to remember what number it follows. When a new pair appears, stop thinking about the last one and concentrate on the new pair. Let's go through the list once this way.

After the first time through the list the following instructions were read to the *S*:

> That's fine. Now we'll go through the list a number of times. Between trials on the list there will be a few blank spaces in the window during which time you can take a short rest, but keep looking at the window. The order of pairs will be different each time through the list, but each number is always followed by the same word. Starting with this next time through the list, try to guess ahead of time what word follows the number you are looking at. In other words, you are to say each word while you are looking at its number. Each pair appears just once each time through the list. Don't hesitate to guess. When you realize that you have made a mistake, say the correct word aloud as it appears in the window. You can expect to make many mistakes, but keep trying and you will learn which word goes with each number.

The *S*s were run for Trials 2–18 without further interruption.

Results and Discussion

The analysis of responses from Exp. II shown in Table 1 was similar to that of Exp. I, save that correct responses were not included in the

TABLE 3

INDIVIDUAL DIFFERENCES IN MEAN NUMBER OF FORWARD AND BACKWARD RESPONSES: EXP. II

Group	Position	Anticipation								
		1	2	3	4	5	6	7	8	9
1–6	Forward	1.42	1.44	1.61	1.73	1.68	1.69	2.0	2.4	
	Backward		.13	.33	.44	.53	.77	.72	.79	.85
7–12	Forward	.94	1.09	1.23	1.42	1.35	1.33	1.05	1.60	
	Backward		.00	.15	.20	.35	.56	.73	.64	.72
13–18	Forward	.55	.57	.67	.80	.75	.50	.58	1.0	
	Backward		.00	.00	.11	.25	.23	.39	.24	.25

analysis. The *S*s of Exp. II tended to respond with words not previously seen on the trial, as is clear from the marked differences in frequency of Forward and Backward errors. Two-tailed binomial tests of the difference in sign between the two kinds of error were significant at the .01 level at each anticipation from A-2 through A-8. The means from beginning to end of a trial in the Forward category tended to increase, although a sign test between A-1 and A-9 was not significant. The increase in Backward errors as a trial progressed is a change from Exp. I. Twenty-eight of the *S*s showed this increase ($p < .01$). It would seem that under the conditions of Exp. II the *S*s' memories did not perform as efficiently as in the previous experiment. The shorter anticipation period coupled with introduction of numbers suggest themselves as relevant factors.

Since it was apparent that there were large differences in learning ability among the *S*s, they were divided into three groups on the basis of their performance on Trials 17–18. Over all *S*s the proportion of correct anticipations on the last two trials was .45. Separate analyses were made of data from *S*s having 1–6, 7–12, or 13–18 correct. Table 3 indicates that the trends within groups are similar. The differences between groups in overall level of erroneous responding reflects the basis on which the three groups were formed.

Table 2 indicates that *S*s made fewer responses as they proceeded through a trial. During Trials 2–9 the mean proportion of omissions was significantly lower for P-1 than for P-9 ($t = 3.93$; $p < .01$). During Trials 10–18 the corresponding means were again different ($t = 3.26$; $p < .01$). Increasing numbers of omissions as a trial progresses apparently is characteristic of list learning in general, and not just serial learning.

The two experiments taken together support the interpretation that *S*s do not, in general, utter the first response that occurs to them, even as a guess, when a list is being tested repeatedly. A recognition mechanism is involved in anticipation which results in the *S*s refraining from speaking responses which they judge to be erroneous. Even with the 2-sec. anticipation interval of Exp. II the *S*s were able to utilize such a mechanism. The paradox arises that the most recent responses, which under other conditions can be shown to be the strongest responses, are the least likely to be spoken.

The results of the present experiments

indicate that the occurrence of anticipations is influenced by guessing strategies as well as by associative strength between verbal items. Number of correct anticipations over a trial gives an inflated estimate of associative strength for which no simple correction seems feasible. Such an inflation may well be related to other verbal learning phenomena such as the differential effects of various degrees of overlearning. Associative strength would be expected to increase as trials continue beyond the point where all items are anticipated correctly.

REFERENCES

BUGELSKI, B. R. A remote association explanation of the relative difficulty of learning nonsense syllables in a serial list. *J. exp. Psychol.*, 1950, **40**, 336–348.

DEESE, J., & KRESSE, F. H. An experimental analysis of the errors in rote serial learning. *J. exp. Psychol.*, 1952, **43**, 199–202.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

PETERSON, L. R., PETERSON, M. J., & MILLER, A. G. Short-term retention and meaningfulness. *Canad. J. Psychol.*, 1961, **15**, 143–147.

PETERSON, L. R., SALTZMAN, D., HILLNER, K., & LAND, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *J. exp. Psychol.*, 1961, **62**, 302–309.

(Received April 18, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 265–269

EFFECT OF DIFFERENT STIMULUS FREQUENCIES ON DISCRIMINATION LEARNING WITH PROBABILISTIC REINFORCEMENT[1]

JULIET POPPER SHAFFER

University of Kansas

A model developed by Burke and Estes generated predictions for a 2-stimulus (T_1, T_2), 2-response (A_1, A_2) successive discrimination learning problem, which consisted of predicting which of 2 reinforcing events, E_1 or E_2, would occur on each of the 324 trials. The probability of E_1 was 1.00 on T_1 trials and .50 on T_2 trials. 3 groups of 32 Ss each differed in the probability (β) of T_1 trials. It was predicted that the final observed proportions of A_1 responses on T_1 and T_2 trials [$P(A_1|T_1)$ and $P(A_1|T_2)$] would be linear increasing functions of β, with equal slopes. The final $P(A_1|T_1)$ was a linear increasing function of β ($p < .01$) while the final $P(A_1|T_2)$ was a linear decreasing function ($p < .05$). Thus, the results did not conform to the predictions derived from the model.

The present paper reports a study of the behavior of human *S*s in a two-stimulus, two-response successive discrimination problem, of the following kind. On each trial, one of two stimuli, T_1 or T_2, is presented, and the *S* predicts which of two reinforcing events, E_1 or E_2, will follow. When T_1 is presented, E_1 has probability π_1 and E_2 has probability $1 - \pi_1$ of occurring. When T_2 is presented, E_1 occurs with probability π_2 and E_2 with probability $1 - \pi_2$. The parameters π_1 and π_2 may take on any values between 0 and 1.

A number of studies of this type of discrimination learning (e.g. Atkinson, Bogartz, & Turner, 1959; Estes & Burke, 1955; Popper & Atkinson, 1958) have been designed, at least in part, as tests of a model developed for this situation by Burke and Estes (1957), based on statistical learning theory. The principal assumption in this model, which is a direct extension of a simple learning model (Estes & Burke, 1953), is that the two stimuli, T_1 and T_2, can be conceptualized as sets of stimulus elements. Every element is assumed to be conditioned to one of the two responses, predicting E_1 or predicting E_2, which will be designated A_1 or A_2, respectively. On T_i trials ($i = 1$ or 2), a sample is taken from the set of elements corresponding to Stimulus T_i. The probability of each response is equal to the proportion of elements in the sample conditioned to it. All elements in the sample on a given trial are assumed to become conditioned to the response corresponding to the reinforcing event on that trial, leading to altered response probabilities on subsequent trials.

Burke and Estes derived theoretical learning curves for both T_1 and T_2 trials. Only the asymptotic predictions will be presented here, since they are the only predictions of relevance to the present study. They are as follows:

$$p_\infty(A_1|T_1) = \pi_1(1 - \omega_1) + \pi_c\omega_1 \quad [1]$$

$$p_\infty(A_1|T_2) = \pi_2(1 - \omega_2) + \pi_c\omega_2 \quad [2]$$

where $p_\infty(A_1|T_i)$ is the expected asymptotic probability of an A_1 re-

---

[1] This research was conducted at Indiana University while the author held a National Science Foundation Postdoctoral Research Fellowship.

sponse given a T_i trial, and π_c is equal to $\beta\pi_1 + (1 - \beta)\pi_2$.[2] The parameters ω_i are functions of the degree of similarity between Stimuli T_1 and T_2, with similarity represented in the model by the assumption that the sets available for sampling on T_1 and T_2 trials, respectively, overlap; i.e., some of the elements available on T_1 trials are also available on T_2 trials. The parameter ω_i is equal to the expected proportion of elements sampled on T_i trials which are common to both sets. Due to the symmetry of the design of the present experiment, it is natural to assume that ω_1 and ω_2 are equal. Finally, β is equal to the probability of a T_1 trial, with T_1 and T_2 trials occurring in random succession.

The studies by Popper and Atkinson (1958) and Atkinson, Bogartz, and Turner (1959) allowed a test of these predictions by using various combinations of π values, keeping other parameters fixed. Both studies found serious discrepancies between observed asymptotic values and the predictions based on the Burke and Estes model. In the present study, π_1, π_2, and the physical stimuli (therefore, presumably, ω) were fixed, and the parameter β was systematically varied. Although the model had not been successful in accounting for the effects of varying values of π, with β fixed, it still appeared possible that it could account for the nature of the changes when β was varied keeping the values of π fixed.

Rewriting Equations 1 and 2 as functions of β, and assuming ω_1 and ω_2 are equal, gives:

$$p_\infty(A_1|T_1) = [\omega(\pi_1 - \pi_2)]\beta + [(1 - \omega)\pi_1 + \omega\pi_2] \quad [3]$$

$$p_\infty(A_1|T_2) = [\omega(\pi_1 - \pi_2)]\beta + \pi_2 \quad [4]$$

from which it can be seen that, if π_1 is greater than π_2, and ω is not equal to zero, both asymptotic proportions should be linear increasing functions of β, with equal slopes. Therefore, their average should also be a linear increasing function of β, and the difference between them should be a constant, independent of β.

METHOD

Subjects.—The *S*s were 96 Indiana University students taking introductory psychology. They were tested in groups of 4 *S*s each.

Apparatus.—A darkened experimental room contained five booths, all facing the front of the room. The *E* sat in the center booth while the 4 *S*s sat in the outer booths. Two 7½-w. bulbs, 25 in. apart, were centered on the front wall, about 10 ft. away from the *S*s. The left bulb was blue, the right bulb yellow. Each booth contained a 1½-in. lever on the right side of a panel, and, 6 in. to the left, two reinforcing lights arranged vertically 6 in. apart, above and below the lever. The lever could be moved up or down, and contained a spring which returned it to center when released. A punched-tape programing device controlled the sequence of stimuli (yellow and blue lights), and the sequence of reinforcing lights in the booths. The *S*s' responses were recorded automatically.

Procedure.—When *S*s had been seated in the booths, they were given instructions to the effect that they were in a prediction experiment which would consist of a series of trials. Each trial would begin with the lighting of a blue or yellow bulb, and shortly afterwards either the upper or lower light on their panel would go on. They were instructed to predict which of the panel lights would go on by moving their levers either up or down on every trial, immediately after the blue or yellow light went on.

Following 4 practice trials, the 324 experimental trials were presented in continuous sequence. On each trial, the stimulus appeared for 2.7 sec., followed immediately by the reinforcing light for 1.8 sec., and there was a 2.0-sec. intertrial interval.

---

[2] These results, in slightly modified form, are presented in Equations 18 and 19 of the Burke and Estes (1957) article.

Design.—For all *S*s, π_1 was equal to 1.00 and π_2 was equal to .50. The three main experimental groups differed in the value of β, the proportion of T_1 trials. For Group I, β was equal to one-third; for Group II, it was equal to one-half; and for Group III it was equal to two-thirds. The design was counterbalanced by having four subgroups, with different identification of the stimuli and responses, within each main experimental group. Two groups of 4 *S*s each were assigned randomly to each subgroup.

Each group of 4 *S*s was given a different sequence of stimuli and reinforcements. The sequences were randomized with the restriction that, within each successive 36-trial block, the numbers of T_1 trials followed by E_1, T_2 trials followed by E_1, and T_2 trials followed by E_2 were exactly equal to their expected values.

RESULTS

The proportions of A_1 responses on T_i trials will be designated $P(A_1|T_i)$. The 324 trials were divided into nine 36-trial blocks, and $P(A_1|T_1)$ and $P(A_1|T_2)$ were computed for each group for each block. The results are shown in Fig. 1. Figure 1 indicates the comparative positions of the groups after equal total numbers of trials, but not after equal numbers of trials of the same type. For example, each point of the $P(A_1|T_1)$ curve for Group I ($\beta = 1/3$) is based on 12 T_1 trials, while each point of the same curve for Group III ($\beta = 2/3$) is based on 24 T_1 trials. To determine the effect of the method of dividing up the trials, the proportions were determined also for blocks of 18 T_1 and 18 T_2 trials, and these are shown in Fig. 2. Figure 2 indicates the comparative positions of the groups after the same number of trials on each stimulus, but after different total numbers of trials. All conditions can be compared up to the sixth block using this scoring method.

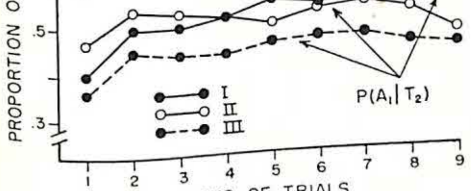


FIG. 1. $P(A_1|T_1)$ and $P(A_1|T_2)$ over 36-trial blocks for Groups I, II, and III.

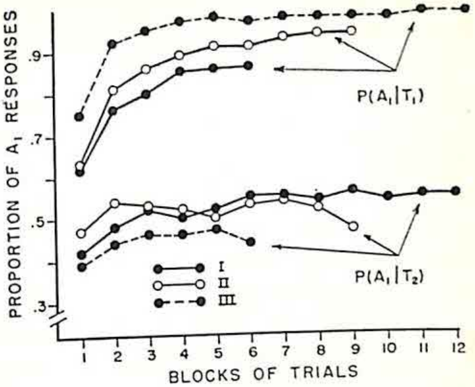


FIG. 2. $P(A_1|T_1)$ and $P(A_1|T_2)$ over blocks of 18 T_1 trials and 18 T_2 trials for Groups I, II, and III.

Since there was no apparent reason for preferring one method over the other, all analyses were carried out on both sets of data, using the proportions from the ninth block in Fig. 1, and the proportions from the sixth block in Fig. 2. With a minor exception, which will be noted below, the conclusions from both sets of analyses were identical. Therefore, only the results using the data from the ninth block in Fig. 1 will be presented in detail. The proportions, $P(A_1|T_1)$ and $P(A_1|T_2)$, for that block are: Group I, .870 and .551; Group II,

TABLE 1

Analysis of Variance of $P(A_1|T_1)$ and $P(A_1|T_2)$ on the Ninth 36-Trial Block

Source	*df*	*MS*	*F*
Group (G)	2	27.69	—
Stimulus-Response identification (I)	3	246.52	—
G × I	6	205.23	—
4-*S* group (4-*S*G)	12	287.89	1.36
S within 4-*S*G (*S*)	72	211.06	—
Trial type (T)	1	96,078.26	366.03**
G × T	2	2,182.16	8.31**
Linear	1	4,220.51	16.08**
Quadratic	1	143.81	—
I × T	3	169.89	—
G × I × T	6	301.70	1.15
4-*S*G × T	12	262.49	—
S × T	72	367.03	—

Note.—Percent scores were used.
** $p < .01$.

.941 and .469; Group III, .992 and .443.

An analysis of variance was performed on these scores, allowing for the effects of experimental group, stimulus-response identification, 4-*S* group, and type of trial (T_1 or T_2). The results are given in Table 1. The effects of trial type and the Group × Trial Type interaction were highly significant, while no other effects approached significance. The Group × Trial Type interaction was further analyzed into linear and quadratic components, and only the linear component was significant. This indicates that the difference between response proportions on T_1 and T_2 trials is a linear function of the value of β, for the values used in this experiment. Since there was no evidence for any differences among the 4-*S* groups, the sum of squares for 4-*S* groups was pooled with the *S* sum of squares in subsequent analyses to achieve greater power.

The difference among groups on T_2 trials only was investigated, using an *F* ratio based only on those trials. The difference approached significance ($F = 2.38$, $p = .10$), and, when analyzed into linear and quadratic components, the linear component was significant ($F = 4.37$, $p < .05$).[3]

The data from the T_1 trials were not suitable for a similar analysis, as the proportions were highly skewed, especially in Group III.[4] A chi square analysis was used, dividing the *S*s in each group into those giving 100% A_1 responses and those not doing so. (Although this test is slightly biased against obtaining significance, since 100% means only 12 A_1 responses in Group I and 24 A_1 responses in Group III, the results using the data from the sixth block in Fig. 2, where no such bias exists, were the same.) The differences were significant ($\chi^2 = 8.67$, $p < .05$). A further chi square test was performed, using data from Groups I and III only. This seemed to be a reasonable equivalent to the analysis of variance test for the significance of the linear component on the T_2 trials, since the linear component for three groups is a function of the difference between the means of the two extreme groups. The difference between Groups I and III was highly significant ($\chi^2 = 8.58$, $p < .01$).

Discussion

The results clearly do not support the use of the Burke and Estes discrimination learning model in interpreting this experiment. Although $P(A_1|T_1)$ on the last block of trials appears to be a linear increasing function of β, as predicted,

---

[3] Using the data from the sixth block in Fig. 2, both the differences among groups and the linear component of the differences were significant ($p < .05$).

[4] This skewness does not affect the overall analysis of variance, as the distributions relevant for that analysis are the distribution of the average of $P(A_1|T_1)$ and $P(A_1|T_2)$ and the distribution of the difference between them.

$P(A_1|T_2)$ is a linear decreasing function, contrary to the prediction, for the values of β used. The two trends are apparently about equal in magnitude, since the average of the two proportions does not differ significantly among groups (as indicated by the nonsignificant group effect in the overall analysis of variance), and, in fact, is almost identical for the three groups (.710 for Group I, .705 for Group II, .717 for Group III). According to the model, the average of the proportions should be a linear increasing function of β. Finally, the difference between the two proportions, which according to the Burke and Estes model should be equal for the three groups, is a highly significant linear increasing function of β, as indicated in the Group × Trial Type interaction effect in the analysis of variance.

It should be pointed out that, since this experiment was designed primarily to test predictions from a model and not as a parametric study, only three values of β were used, and therefore many types of departure from linearity could not be detected. Nevertheless, within these limits, the systematic way in which the data deviate from the predictions should encourage further theoretical analysis of this discrimination learning situation.

REFERENCES

ATKINSON, R. C., BOGARTZ, W. H., & TURNER, R. N. Supplementary report: Discrimination learning with probabilistic reinforcement schedules. *J. exp. Psychol.*, 1959, **57**, 349–350.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

ESTES, W. K., & BURKE, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.

POPPER, J., & ATKINSON, R. C. Discrimination learning in a verbal conditioning situation. *J. exp. Psychol.*, 1958, **56**, 21–25.

(Received April 18, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 270–279

EFFECT OF AN ANCHOR STIMULUS ON THE STIMULUS GENERALIZATION GRADIENT[1]

LEO GANZ[2]

University of Chicago

Rhesus macaques were trained to press a key differentially to 2 monochromatic hues 150 mμ apart in wave length. The stimuli were presented through contact lens diffusers. S^d was presented under a 15-sec. variable interval schedule of positive reinforcement (5% sucrose solution) and S^Δ was presented under a 15-sec. delay-of-reinforcement schedule. Generalization was measured to the intervening hues, covered in 20-mμ steps. The generalization testing was preceded by 1 hr. of stimulus exposure and a 10-min. period of dark adaptation. During stimulus exposure, Ss sat in the primate chair, head restrained, wearing the contact lens diffusers, and exposed to a repeating cycle of hues. It was found that when stimulus exposure consisted of colors within 20 mμ above and below S^Δ, response rate during generalization testing was consistently enhanced at S^d.

Generalization refers to the tendency of organisms to maintain response levels when a training stimulus is replaced by a *similar* stimulus. Psychophysical studies (e.g., Helson, 1947) have shown that the context in which a stimulus is presented can modify its similarity, as estimated verbally by human *S*s. Consider this example of successive contrast: two stimuli are verbally estimated to be more similar to one another when a third stimulus, an extreme value, the so-called anchor, is presented. If the generalization gradient is a reflection of stimulus similarity, it is reasonable to expect that it can be modified when the generalization stimuli are presented in such differing contexts.

Successive contrast has been measured with a behavioral index by Pavlov and his colleagues (1927) in experiments on positive and negative induction. In positive induction, the repeated application of CS^- led to an *increase* in the response to CS^+. Gewirtz, Jones, and Waerneryd (1956) have shown sequential effects of exposure to one series of stimuli upon the generalization gradient elicited by a subsequent series. The gradient was steeper on the side of the stimulus dimension close to the previous series and the gradient mode or peak was shifted away from this previous series. Hanson (1959) and Pierrel and Sherman (1960) have reported generalization gradients whose peaks are displaced from S^d in a direction away from S^Δ. This peak shift is more pronounced the closer S^d and S^Δ are on the continuum. In the studies of Reynolds (1961a, 1961b) and Reynolds and Catania (1961), the introduction of S^Δ has been found to induce a rise in S^d responding, a phenomenon they term *behavioral contrast.*

The present study continues the analysis of the effects of stimulus exposure on the generalization gradient. Specifically, the range and locus of

---

[1] This study was supported in part by Research Grants B-771 and B-1590 from the National Institutes of Neurological Diseases and Blindness of the National Institutes of Health, United States Public Health Service. The author wishes to thank Austin H. Riesen for his advice and assistance throughout the course of the investigation.

[2] Now at University of California, Riverside.

this stimulation are varied relative to the training stimuli. In two series of experiments, rhesus macaques are trained to discriminate a pair of monochromatic stimuli. The *S*s are then exposed to successive series of cycled monochromatic stimuli under extinction, but not in a discrimination setting. They are tested for generalization following such stimulation.

GENERAL METHOD

Subjects.—Six preadolescent *Macaca mulatta* and *Macaca radiata*—Fo, Ha, Ch, Ka, Aj, Na, and Bu—served as experimental *S*s. These animals had acquired a number of previous wave-length discriminations (see Ganz & Riesen, 1962).

Conditions of rearing.—During the course of the experiment, visual stimulation outside the experimental situation was excluded by rearing the *S*s in the laboratory darkroom. (See Ganz & Riesen, 1962, for further details of rearing.)

The *S*s were maintained on a 22-hr. hunger-thirst drive. Their diet was comprised of approximately 100 gm. Purina monkey chow per kg. body weight immersed in water. Key pressing was the behavioral index under analysis. Reinforcement consisted of 3–4 drops of tube-fed, 5% sucrose solution.

Apparatus.—Two similar primate chairs were used: one for stimulus exposure sessions and one for training and testing sessions. These chairs allowed control over the proximal stimulus, i.e., the stimulus as it impinges on the receptor surface. The training and testing chair was inside a light-tight cubicle (Ganz & Riesen, 1962). The chair included arrangements for restricting the animal's head and a neckboard which prevented the animal from touching his face; these restrictions maintained him in the chair, but allowed a wide variety of movement by the rest of the body. A beam of monochromatic illumination impinged on the animal's left eye. The training and testing chair differed only in that it contained, in addition, a manipulandum—a key-operated microswitch—which was located in front of *S* and under the neckboard. When a reinforcement was available, a key press activated a solenoid valve for 1 sec. This released 3–4 drops of 5% sucrose solution, directed to the animal's mouth via a tube.

During the experimental session, *S*s wore contact lenses over their eyes (Ganz & Riesen, 1962). These diffusers were introduced to prevent *S* from exercising any receptor orienting or "observing" responses.

A set of condensing lenses focused the filament of a 300-w. projection lamp onto the aperture of the cubicle. This was in turn focused as a ¾-in. diameter disc on the surface of the contact lens. Bausch and Lomb interference filters Series 33–78 were interposed in this optical path to select 8-mμ half-band-widths of monochromatic light from 449 to 631 mμ, in 10-mμ steps. In addition, a Kodak Wratten gelatin filter No. 8 absorbed third-order interference peaks in the violet and blue regions present with filters of peak transmission above 500 mμ. Polarizers were partially crossed and interposed in the beam to approximate an equal luminosity spectrum for the human. A Macbeth illuminometer was used in a heterochromatic reading of the illumination at the surface of the white contact lens. The 567-mμ stimulus projected 12.4 ft-c of illumination.

These filters were maintained on a wheel which could be made to present a sequence of colors automatically. White noise was fed to a 12-in. speaker above *S*'s head. This further masked cues from the sequencing and recording apparatus outside the cubicle, itself sound-insulated.

General daily procedure.—Training and testing was administered 6 days a week. The *S*s alternated in their daily order. Each *S* was given first a 60-min. stimulus exposure session in one of the primate chairs, within one cubicle and then a 30-min. training and testing session in the primate chair in the second cubicle.

The *S* was carried from his darkroom cage to the *stimulus exposure chair* in a light-proof box. In the experimental cubicle he was placed in the chair, the two contact lenses, one opaque over the right eye and a diffuser over the left eye. A shutter was then opened permitting the stimulus beam of light to illuminate the contact lens. The *E* then left the cubicle and set the control apparatus in operation. A slowly rotating wheel containing four interference filters now exposed the animals to repetitive cycles of colors, varying in range, for 60-min. periods. Each of the stimuli came on for 2 sec. with 1-sec. intervals between stimuli. At the end, *S* was transferred via light-proof box to the second primate chair and similarly installed, again in the absence of light. The next *S* was now installed in the first chair. It required approximately 10 min. to complete the transfer of *S*s. Since *S*s were deprived of light during

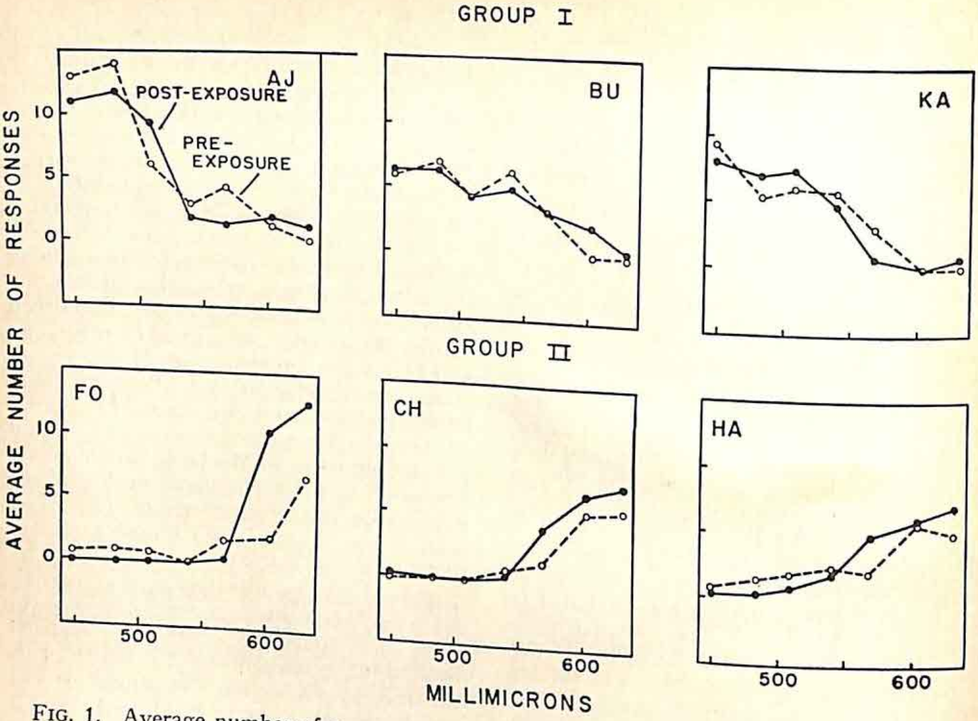


FIG. 1. Average number of responses per stimulus presentation emitted by individual *S*s as a function of the wave length of the generalization stimulus: Exp. I.

transfer, the transfer period constituted a 10-min. dark adaptation period.

In the second chair, *the testing chair*, S^d and S^{Δ} alternated in irregular sequence. The stimulus-on period was 15 sec. and the -off period was 5 sec. Key presses during S^d yielded reinforcement on a 15-sec. VI schedule. During the stimulus-off period and during S^{Δ}, key presses prolonged S^{Δ} for 15 sec.

Design.—A series of seven (Exp. I) and eight (Exp. II) generalization stimuli (GSs) covered the S^d to S^{Δ} range in approximately equal steps. One such series was presented daily. Order of presentation within a series was predetermined by a balanced design. A series of seven or eight GSs on a testing day was broken down into an initial four GSs, given immediately after *S* was seated in the *testing chair* and another three or four GSs given after 13 training trials had intervened. This was followed by sufficient training trials to complete a 30-min. session. The GSs were administered on alternate days under extinction and under a 14-sec. VI schedule of positive reinforcement.

EXPERIMENT I

Method

The 6 *S*s were divided into two groups. Group I comprised the 3 *S*s (Aj, Bu, Ka) who formerly had had discrimination training involving the following stimuli and values, in millimicrons wave length: $S^d = 449$ vs. $S^{\Delta} = 509$; $S^d = 567$ vs. $S^{\Delta} = 509$; $S^d = 567$ vs. $S^{\Delta} = 631$. This group was trained with the values: $S^d = 449$ vs. $S^{\Delta} = 631$. Group II comprised 3 *S*s (Fo, Ch, Ha) who had had discrimination training involving the following stimuli and values: $S^d = 509$ vs. $S^{\Delta} = 449$; $S^d = 509$ vs. $S^{\Delta} = 567$; $S^d = 631$ vs. $S^{\Delta} = 567$. This group was now trained with the values $S^d = 631$ vs. $S^{\Delta} = 449$.

When discrimination training was begun, all animals were, because of their former training, already responding to S^d, and at a lower rate to S^{Δ} and to the blackout period. Training sessions were continued until criterion level was attained. The criterion of discrimination learning was met if more than 100 responses were emitted during a session and more than 80% of them during S^d, over 2 consecutive days.

During generalization testing, eight successive runs were administered over a period of 8 days. A run was made up of 1 hr. of *stimulus exposure* in the exposure chair, then 10 min. of dark adaptation during the transfer operation, and then a 30-min. *test session.* The *stimulus exposure* hour was composed as follows: during the first 3 days, no stimuli were presented during the stimulus exposure hour, i.e., the *S*s were in darkness; during the last 5 days, the stimuli 443, 457, 469, 489 mμ were cycled, repetitively, and in that order for 1 hr. The test session was made up of the following progression of stimuli: four training stimuli S^d and S^Δ; four generalization stimuli in the order 567, 509, 449, and 604 mμ presented under extinction; 13 training trials with S^d and S^Δ; the three generalization stimuli 483, 631, and 540 mμ; retraining to complete a 30-min. session.

Results

The gradients were collected adding each *S*'s total number of responses emitted in the presence of each generalization stimulus over the first 3 test days (blackout stimulus exposure) and over the last 5 days (blue stimulus exposure). The gradients were then made comparable by dividing the totals by the number of test days.

These gradients are depicted in Fig. 1. Group I, whose S^d adjoined the blue colors of the exposure hour, did not show any consistent effect. Group II, whose S^Δ adjoined the colors of the exposure hour showed consistent results in the 3 *S*s: there is a depression in response rate near S^Δ and an enhancement in response rate near S^d.

EXPERIMENT II

The first experiment had suggested that the effect of stimulus exposure on the generalization gradient depended on the location of the exposure stimuli relative to the S^d and the S^Δ. The position of the stimulus exposure series on the stimulus continuum was the parameter of the second experiment.

Method

The stimulus exposure cycle for both Groups I and II was located as follows: near S^d, near S^Δ, covering the middle range between S^d and S^Δ, covering the whole range from S^d through and including S^Δ, and placed at a distance from the S^d to S^Δ range (anchor condition), in addition to a no-exposure condition; viz., darkness.

The same *S*s were used, with Bu excluded midway in the series because of illness. Group I, which had been trained to discriminate $S^d = 449$ mμ from $S^\Delta = 631$ mμ, was now transferred to $S^d = 449$ mμ vs. $S^\Delta = 610$ mμ; Group II which had been trained to discriminate $S^d = 631$ mμ from $S^\Delta = 449$ mμ, was now transferred to $S^d = 610$ mμ vs. $S^\Delta = 449$ mμ. This transfer occurred readily in all cases and little additional training was required to bring the *S*s to their former criterion level.

Each day, a 1-hr. stimulus exposure period was administered, then a 10-min. dark adaptation period during chair transfer, and

TABLE 1

SUCCESSION OF HUES DURING THE STIMULUS EXPOSURE CYCLES

Cycle	Hue	Order of Presentation						
		1	2	3	4	5	6	Etc.
1	No	No Visual Stimulation						
2	Blue	443[a]	457	483	503	443	457	. . .
3	Red	631	625	610	588	631	625	. . .
4	Yellow-Green	519	540	564	580	519	540	. . .
5	Broad range	625	580	540	483	625	580	. . .
6	Anchor	740	700	660	636	740	700	. . .

[a] Peak transmission of monochromatic filters in mμ. See text for filter characteristics.

then a 30-min. testing and training session, exactly as in Exp. I. A stimulus exposure cycle was presented over 8 successive days. The entire experiment consisted of six types of cycles, 8 days each, to produce a total of 48 successive days of exposures and generalization testing. The manner in which the cycles were composed and ordered is shown in Table 1 for both groups. These stimuli were cycled as in Exp. I.

Following the exposure hour and during the 10-min. dark adaptation period, *S* was transferred to the testing chair. The *S* was then given four generalization tests, then 13 training trials, another four tests, and finally sufficient retraining to complete a 30-min. session. For all *S*s, the following balanced orders of presentation were used in presenting the GS (in mμ): Days 1 and 2: 527, 469, 546, 588, retraining, 567, 489, 610, 509, retraining; Days 3 and 4: 588, 546, 469, 527, retraining, 509, 610, 489, 567, retraining; Days 5 and 6: 567, 489, 610, 509, retraining, 527, 469, 546, 588, retraining; Days 7 and 8: 509, 610, 489, 567, retraining, 588, 546, 469, 527, retraining. Each of these orders was used on 2 succeeding days, Days 1, 3, 5, and 7 with the GSs under extinction, and Days 2, 4, 6, and 8 with all the GSs under a 15-sec. VI schedule of reinforcement. One change was introduced during the sixth cycle, the "anchor" stimulus exposure: a 30-min. dark adaptation period intervened between the stimulus exposure hour and the testing.

Results

Gradients of stimulus generalization were again collected by adding the total responses of each *S* in the presence of each GS, summed over 8 days within each cycle of stimulus exposure. These gradients were subjected to an analysis of variance with one between-group comparison (Group I vs. Group II) and two within-*S*s comparisons (exposure cycle and stimuli). The results are given in Table 2. Any effect on gradient slope of the stimulus exposure treatment common to both groups would contribute to an Exposure × Stimuli interaction of sum of squares (*SS*), which in this case proved significant, $F\ (35,\ 105) = 2.11,\ p < .01$. Any effect of the stimulus exposure treatment different in its direction from Group I to Group II would contribute to the Exposure × Stimuli × Groups interaction *SS*. This term also proved to be significantly above chance levels, $F\ (35,\ 105) = 1.82,\ p < .05$. It can be concluded that the cycles of stimulus exposure had different effects on the stimulus generalization gradient, from cycle to cycle, depending on the composition of the cycle, and from group to group, depending on the position of the S^d and S^Δ.

TABLE 2

ANALYSIS OF VARIANCE OF GENERALIZATION GRADIENTS FOLLOWING STIMULUS EXPOSURE

Source	*df*	*MS*	*F*
Groups (G)	1	10,840.6	3.65
Exposure (E)	5	380.9	<1.00
Stimuli (S)	7	642.9	<1.00
G × E	5	1,334.2	2.80
G × S	7	10,602.4	9.32**
E × S	35	133.8	2.11**
G × E × S	35	115.4	1.82*
Error (b)	3	2,952.8	
Error $(w)_1$	15	476.7	
Error $(w)_2$	21	1,137.3	
Error $(w)_3$	105	63.5	

* $p < .05$.
** $p < .01$.

TABLE 3

INTERACTION ANALYSES OF VARIANCE OF GENERALIZATION FOLLOWING STIMULUS EXPOSURE

Exposure Cycles Paired	*MS*	*F*
1 vs. 2	170.8	2.70*
1 vs. 3	28.9	<1.00
1 vs. 4	40.8	<1.00
1 vs. 5	27.2	<1.00
1 vs. 6	392.0	6.17**
2 vs. 3	190.0	3.06**
3 vs. 4	48.3	<1.00
4 vs. 5	22.0	<1.00
5 vs. 6	346.0	5.44**

Note.—Error $(w)_3$ term from analysis on total data was used to compute the *F* ratios (see Table 2).
* $p < .05$.
** $p < .01$.

We next analyze which cycles

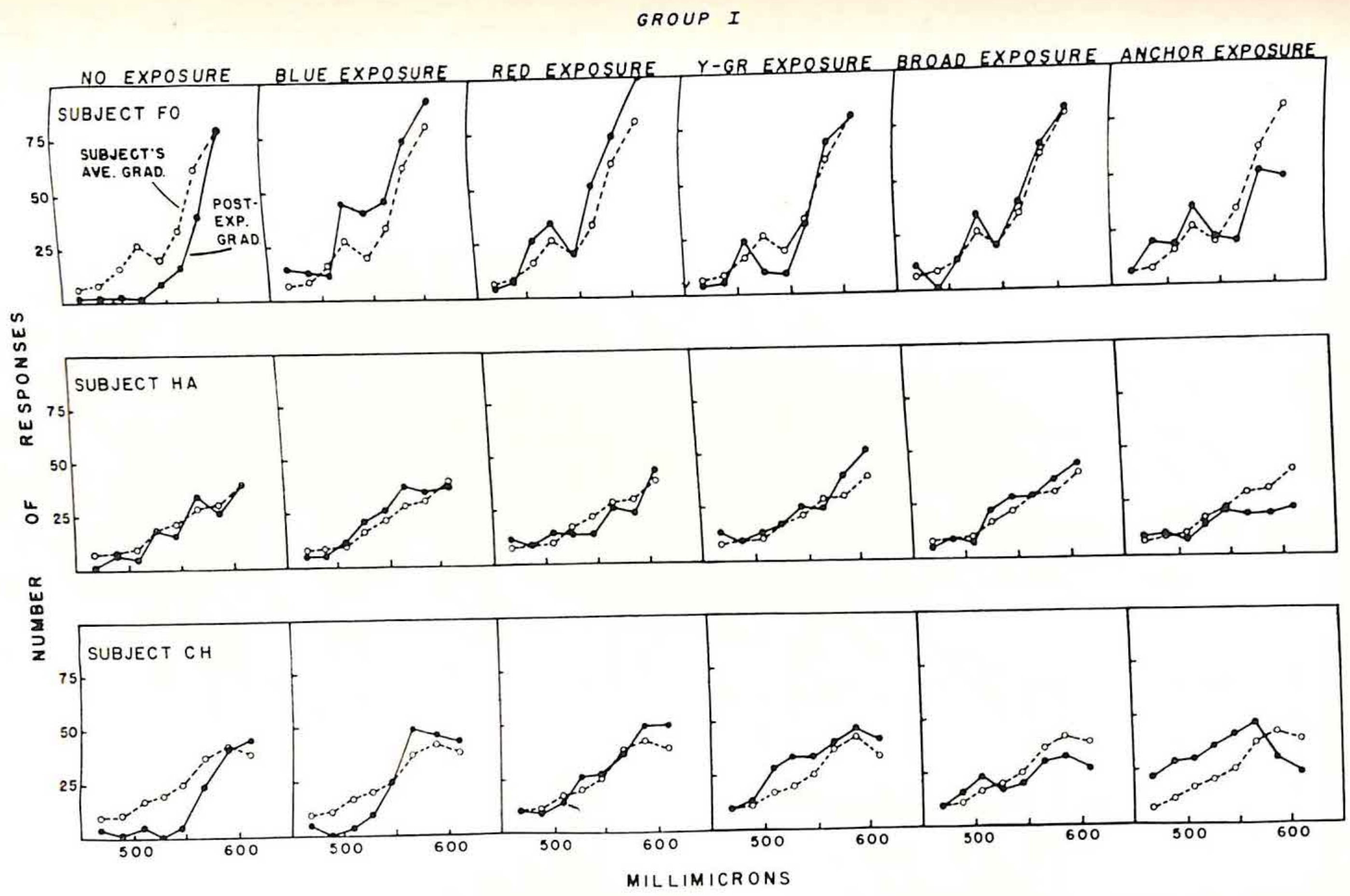


FIG. 2. Total responses emitted by individual *S*s in Group I as a function of the wave length of the generalization stimulus. (The parameter is the hues which comprise the preceding stimulus exposure.)

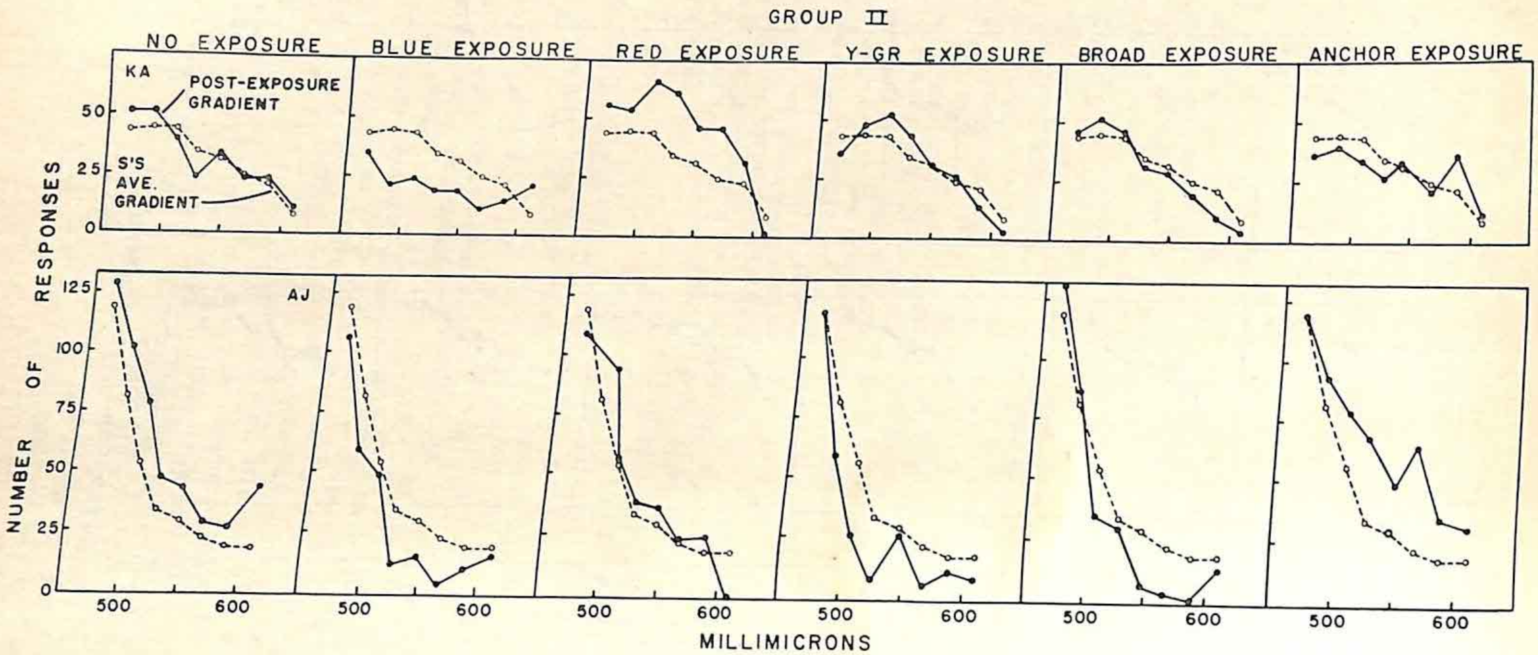


FIG. 3. Total responses emitted by individual *S*s in Group II as a function of the wave length of the generalization stimulus. (The parameter is the hues which comprise the preceding stimulus exposure.)

modified the gradient. We do this by examining the Exposure × Stimuli interaction term for individual pairs of exposure cycles. This is shown in Table 3. The upper part of the table depicts the interactions of each cycle of stimulus exposure with Cycle 1, no exposure, presumably a neutral stimulus exposure with respect to a hue continuum. The lower part depicts the *SS* resulting from comparisons of successive pairs of cycles. The comparison of successive pairs provides some control over the effect of order of testing per se. From Table 3, it is clear that the stimulus exposures were not equally effective in modifying the gradient of generalization. The significant treatments seemed to come from Cycles 2, 3, and 6, where the exposure stimuli were near S^d, near S^Δ, or entirely outside the S^d to S^Δ range. The ineffective cycles involved stimuli placed symmetrically between S^d and S^Δ.

The generalization gradients of individual *S*s are depicted in Fig. 2 and 3. Considering Group I first, we see that in Cycle 1 (no exposure) *S*s give crudely bell-shaped gradients from the 610-mμ S^d in the orange to the 469-mμ S^Δ in the blue. An average gradient (dashed line) from all cycles combined is also shown for comparison. In Cycle 2, *S*s were repeatedly exposed to stimuli in the blue region, i.e., a small range of stimuli in the S^Δ locus prior to generalization testing. The effect of such exposure is to raise response rate near S^d. Compare, for example, the post-exposure gradients of Cycle 1 and Cycle 2. This replicates, in the same *S*s, the findings of Exp. I. During Cycle 3, *S*s were exposed to stimuli in the red region, i.e., a small range of stimuli in the S^d locus. The effect is not consistent. Cycles 4 and 5 effects were not statistically significant. Cycle 6 comprised hues far into the red region. The stimuli did not overlap the S^d to S^Δ range. The effect is clearly a depression in response rate at S^d and some rise in rate at S^Δ. Hence in Cycle 6 the gradient is shallower.

Turning to Group II, represented in Fig. 3, we see that Cycle 2, in the blue region but now at the S^d locus, depresses response rate near S^d. It will be recalled that the same blue exposure near S^Δ in Group I had an enhancing effect on response rate at S^d.

Cycle 3 involved stimulus exposure

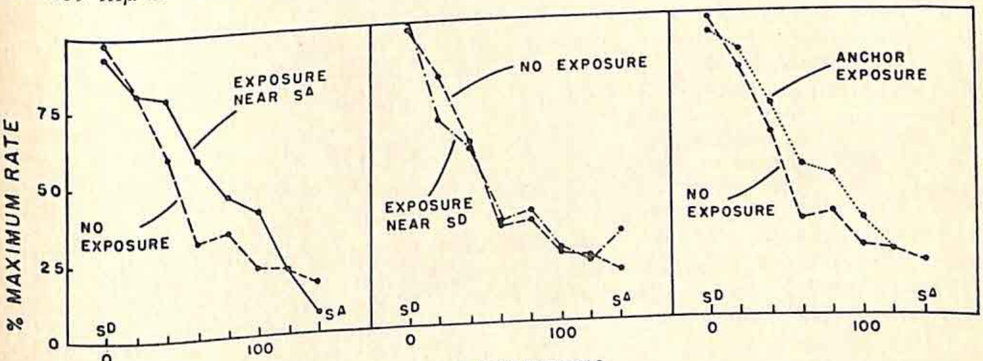


FIG. 4. Relative generalization as a function of wave-length distance from S^d and the position of the preceding stimulus exposure session. (Average taken over the relative gradients of the 5 *S*s.)

in the red region, near S^Δ for Group II. Both *S*s changed their gradients in the direction of enhanced responding near S^d (compare Cycle 2). Response rate at S^Δ was reduced. This result is in many respects similar to the Cycle 2 findings for Group I and to the Exp. I results for Group II. In all these cases, (*a*) the exposure stimuli were in the S^Δ region of the spectrum; (*b*) response rate was enhanced at S^d; (*c*) response rate was depressed at S^Δ; (*d*) generalization was steeper between S^d and S^Δ. The combined relative gradients of *S*s of Groups I and II are shown in Fig. 4.

Cycles 3 and 4 did not contribute significant interaction *SS*s and so will not be discussed. In Cycle 6, stimulus exposure was completely outside the S^d to S^Δ range, but nearer to S^Δ for Group II. One *S* showed a drop in response rate and little effect at S^Δ. The other *S* showed a rise in response rate near S^Δ and little effect at S^d. The gradients were shallower for both *S*s. The *S*s in Groups I and II seemed to have generalized with a more shallow gradient following Cycle 6, the anchor exposure (see Fig. 4).

DISCUSSION

This study reports effects on the generalization gradients not unlike those which have been reported in psychophysical experiments. They, along with other recent findings (Honig, Thomas, & Guttman, 1959; Reynolds, 1961a, 1961b) suggest that the excitatory and inhibitory potentials a stimulus evokes are not independent of the stimulus context in which that stimulus appears. The studies cited in the introduction (Reynolds, 1961a) have shown that exposure to stimuli in the S^Δ region, under extinction, can have the paradoxical effect of enhancing response strength at and around the generalization gradient near S^d. Since the gradient from S^d to S^Δ is thus steeper, we are justified in saying S^d and S^Δ are functionally more dissimilar in their control of behavior. It is in this sense that we speak of successive contrast.

A survey of traditional mechanisms shows they are not sufficient to account for the phenomenon:

1. In the customary discrimination situation, the probability that the next stimulus is an S^d is positively correlated with the length of the run of immediately preceding S^Δs. Thus, length of S^Δ run could, itself, acquire cue properties. This would enhance responding following exposure to many S^Δs. It would not enhance response strength as selectively as has appeared here. Furthermore, such an S^Δ exposure was present in the Cycle 1 condition, the blackout control where no visual stimulus was presented. Reynolds (1961a) has found contrast effects following such a blackout. But the contrast phenomenon described here is in addition to any following blackout.

2. Selective receptor adaptation: the presentation of an S^Δ such as hue desaturates S^Δ and increases S^d saturation if they are sufficiently separated. This might increase S^d response strength. It is precluded here because recovery from light adaptation was allowed to reach asymptote.

3. Is contrast simply a manifestation of *reflex reserve*? Responses inhibited by the control exerted by S^Δ might be emitted during S^d. Thus, a long run of S^Δ presentations would pile up responses during the immediately following S^d. Reynolds (1961a) has shown experimentally that this is not the critical variable in behavioral contrast. It is excluded here for the reasons already considered in Section 1 above.

4. What are the reinforcement effects of preceding generalization testing by exposure to a range of stimuli? Most likely they are inhibitory. During stimulus exposure, *S* is shown stimuli under extinction. There is no manipulandum present, but some mid-air key presses probably occur. In fact, on a number of occasions, *E* joined *S* in the stimulus exposure cubicle and observed such

abortive key presses. Since no reinforcement followed, negative habit strength could well have accrued at the exposure stimuli. Suppose positive and negative habit strength generalizes and sums, roughly as Hull (1943) has described it. Accrual of additional negative habit strength as S^{Δ} would generalize to S^d and, if anything, reduce total effective response strength at S^d. It would obviously *never enhance it*. Thus, the summation model of generalization behavior cannot account for the successive contrast phenomenon (Reynolds, 1961b).

The effect of extinction is, nevertheless, present. It is *superimposed* on the successive contrast phenomenon. Consider the Group II results, the Blue Exposure in Fig. 3; here a lengthy stimulus exposure at S^d depresses response rate at S^d. The same occurs in Group I, where the Anchor Exposure is near S^d. But there is no contrast effect at S^d. Then, in Cycles 4 and 5, where the exposure is between S^d and S^{Δ}, we find occasional drops in response strength between S^d and S^{Δ}. Again, however, there is no contrast effect at S^d. Finally, when the stimulus exposure is at S^{Δ}, there are occasional drops in response strength at S^{Δ}. This condition also enhances response strength at S^d. Thus, as expected, inhibitory potential appears maximal at the locus of the stimulus exposure. But, contrast is only present when the exposure and test stimuli are sufficiently separated. The two effects sum, but must be independent.

It is suggested here that the stimulus generalization gradient is sensitive to a successive contrast mechanism, (e.g., Helson, 1947). The effective position of a stimulus on a continuum is not absolute, but is relative to the population of stimuli currently sampled. For example, Reynolds (1961a) has concluded such a contrast phenomenon determines the potency of the reinforcing stimulus, S^r: "The frequency of reinforcement in the presence of a given stimulus, *relative to the frequency during all of the stimuli that successively control an organism's behavior*, in part determines the rate of responding that the given stimulus controls [Reynolds' italics]" (p. 70). Our results suggest that S^d and S^{Δ} are similarly prone to successive contrast. The presentation of an anchor stimulus appears to displace the apparent location of a test stimulus away from that anchor. So, where the anchor exposure is at S^{Δ}, it makes the S^d function as if it were further from S^{Δ}. Because of the peak shift (Hanson, 1959), response strength is enhanced at the displaced S^d. Such a contrast effect is dependent on the location of the exposure stimulus on the S^d to S^{Δ} continuum, just as the effectiveness of an anchor is greater when it is sufficiently separated from the test stimulus.

REFERENCES

GANZ, L., & RIESEN, A. H. Stimulus generalization to hue in the dark-reared Macaque. *J. comp. physiol. Psychol.*, 1962, **55**, 92–99.

GEWIRTZ, J. L., JONES, L. V., & WAERNERYD, K. Stimulus units and range of experienced stimuli as determinants of generalization-discrimination gradients. *J. exp. Psychol.*, 1956, **52**, 51–57.

HANSON, H. M. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58**, 321–334.

HELSON, H., Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HONIG, W. K., THOMAS, D. R., & GUTTMAN, N. Differential effects of continuous extinction and discrimination training on the generalization gradient. *J. exp. Psychol.*, 1959, **58**, 145–152.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

PAVLOV, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

PIERREL, R., & SHERMAN, J. G. Generalization of auditory intensity following discrimination training. *J. exp. Anal. Behav.*, 1960, **3**, 313–322.

REYNOLDS, G. S. Behavioral contrast. *J. exp. Anal. Behav.*, 1961, **4**, 57–71. (a)

REYNOLDS, G. S. Contrast, generalization, and the process of discrimination. *J. exp. Anal. Behav.*, 1961, **4**, 289–294. (b)

REYNOLDS, G. S., & CATANIA, A. C. Behavioral contrast with fixed-interval and low-rate reinforcement. *J. exp. Anal. Behav.*, 1961, **4**, 387–391.

(Received April 20, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 280–291

DISCRIMINATION AND STIMULUS GENERALIZATION OF APPROACH, OF AVOIDANCE, AND OF APPROACH AND AVOIDANCE DURING CONFLICT[1]

HOWARD S. HOFFMAN AND MORTON FLESHLER

Pennsylvania State University

1 group of rats was trained on approach (plate press to obtain food), a 2nd group was trained on avoidance (bar press to prevent shock), and a 3rd group was trained on both responses. For all 3 groups the discriminative stimulus was a 3500-cps tone. During subsequent tests for stimulus generalization, it was found that approach generalized more broadly than avoidance, but Ss learned to discriminate more rapidly for approach than for avoidance. In the tests during approach-avoidance conflict (Group 3), displacement was observed despite the fact that the gradient of approach was higher everywhere than the gradient of avoidance. While the results conform to certain features of Miller's theory of displacement, they also indicate the need for revisions. In particular, they suggest that conditioned suppression plays an important role in determining behavior during conflict.

There is today a sizeable literature in the area of stimulus generalization and also in the area of discrimination (Kimble, 1961; Mednick & Freedman, 1960). However, as yet there have been no studies which have examined and compared both of these phenomena when the relevant behavior was under aversive, as opposed to appetitive, control. One purpose of the present research was to examine the gradient of stimulus generalization and the subsequent acquisition of an absolute discrimination when the relevant response permitted the *S* to obtain food (approach) and when the relevant response permitted the *S* to prevent electrical shock (avoid). A second purpose was to examine these processes when the two responses (approach and avoidance) were under the control of a single stimulus and hence in conflict.

In addition to providing information about discrimination learning under aversive, as opposed to nonaversive controls, these experiments provide an arena for assessing Miller's (1959) theoretical treatment of psychoanalytic displacement. According to the theory, a response may be displaced to some other stimulus, if the original stimulus exhibits a strong tendency to also elicit an incompatible response, and if the gradient of generalization for the incompatible response falls more rapidly than does the generalization gradient for the original response. Under these conditions the original stimulus may elicit the incompatible response, whereas a stimulus on the wing of the generalization gradient may elicit the response that initially was appropriate to the original stimulus. In Miller's terms, the response is displaced to a new stimulus.

A considerable body of research by Miller and his associates (Miller, 1959) has provided empirical confirmation of the major theoretical predictions and has also yielded support for several of the assumptions upon which the theory is based. However, in none of the research to date, has it been possible to assess the gradients for both the original and incompatible responses during the conflict phase of the work. The present sequence of studies was designed to provide this information.

---

[1] This research was supported by National Institute of Mental Health Grant M-2433.

GENERAL METHOD

Subjects.—The *S*s were 25 female rats of Sprague-Dawley stock. They were approximately 90 days old at the start of the research.

Apparatus.—The experimental chamber was a sound insulated Skinner box fitted with a one-way vision observation port. In the middle of the front wall was a food well to which reinforcement, in the form of Noyes food pellets, could be delivered. On either side of the well was a manipulandum. To the right was a bar of .25 × .75 in. aluminum which projected 1.5 in. into the test chamber at a height of 1.5 in. and which actuated a microswitch when a downward force of 20 gm. was applied. At the left, 1.5 in. above the grid floor, was a plate (1.5 in. square) which protruded .5 in. from, and was parallel to, the front wall. When a horizontal force of at least 15 gm. was applied perpendicular to the plate, it actuated a second microswitch. The circuitry was such that a response (either bar press or plate press) was defined as the initial closure of the corresponding microswitch. Thus, holding responses had no effect on the program.

The aversive stimulus, electrical shock, was generated on a Foringer unit. This device incorporates a 250,000-ohm resistance in series with the animal. The walls, the food well, the manipulanda, and the grid floor were wired to carry the shock and during shock, the polarity of the grid bars, the manipulanda, the food well, and the walls was continuously scrambled so as to make unauthorized escape highly improbable.

In all experiments the discriminative stimuli consisted of pure tones which were programed through a Magnecord PT 6 A tape recorder. By tape recording the acoustic signals, it was possible to arrange that all *S*s received identical stimuli during the several phases of their training and testing and also to arrange the tones (through appropriate splicing techniques) so that their onsets were free of transients. A system of timers, steppers, and relays was used to establish the several contingencies between stimuli, plate presses, bar presses, electrical shock, and food reinforcement. The action of this system was coordinated with the prerecorded sequence of acoustic signals by means of a sensitive relay which was actuated by the electrical output of the tape recorder. Acoustic signals were delivered to the *S* through a 5-in. speaker mounted in the back wall of the chamber. Out of a second speaker, mounted at the side, unfiltered white noise was constantly presented at 60 db. SPL so as to mask sounds having external origin. All of the acoustic signals used in the present research were equated for intensity. To do so, they were recorded on magnetic tape at a level which, during playback, yielded a reading of 80 db. SPL on the C (flat) scale of a General Radio sound level meter when the meter microphone was placed in the center of the experimental chamber and the masking noise was on.

An Esterline-Angus 20-channel operations recorder documented all stimuli, bar presses, plate presses, food reinforcements, and shocks. A Standard Electric timer was used to record latencies.

EXPERIMENT I: ASSESSMENT OF GRADIENTS OF APPROACH AND OF AVOIDANCE

Method

Training on approach.—Five *S*s, maintained at 80% of their previously determined free feeding body weight, and deprived of food for 24 hr., were trained to press the plate for food in the presence of a 3500-cps tone. Each tone lasted 5 sec. and the programing circuit was arranged so that only the first response to a given tone could yield reinforcement. The interval between tones was variable, ranging from 20 to 50 sec. with a mean of 35 sec. In order to sharpen the discrimination between the presence and the absence of tone, the programing circuit was also arranged so that responses in silence, which occurred within 5 sec. of the scheduled onset of a tone, prevented the occurrence of that tone.

When plate press to tone was well developed, the percentage of tones yielding reinforcement was gradually reduced from 100% to 40%. On each session, 70 tones were presented and for each *S*, training was continued until a performance was exhibited in which over two consecutive sessions, responses occurred to at least 98% of the total number of tones and responses in silence were negligible (less than three per session). The attainment of this behavioral requirement involved an average of 15 sessions.

Testing on approach.—Because the frequency response of the tape recorder dropped off sharply above 4000 cps, it was necessary to restrict the tests for stimulus generalization and discrimination to tones having frequencies at 3500 cps and below and hence to limit the tests to one side of the gradient. In an effort to equate roughly sensory differences among stimuli, the frequencies of the tones employed in the tests were spaced approximately one log unit apart. These tones had the following frequencies: 3500, 2280, 1500, 1000, 670, and 400 cps.[2]

During testing, the tones were presented singly and in random order with the following restrictions: each series of stimuli contained three tones at 3500 cps (the S_D employed in training) in addition to each of the five new test frequencies. Within each series of eight stimuli, two of the three 3500-cps tones were programed to yield food reinforcement. During the tests, the tones were presented at intervals ranging from 30 to 50 sec. Test sessions were run daily. In each test session, the sequence of eight tones appeared 12 times, each time in a new random order. Testing was run for a minimum of 10 sessions and was terminated when the *S* exhibited a near perfect discrimination between the tone used in training (3500 cps) and the other test frequencies. The criterion for terminating the test process was a session in which the *S* responded no more than once to the 2280-cps tone (i.e., the tone closest to the S_D) while maintaining the response to the S_D.

Tests were conducted in this fashion for several reasons. First, the occurrence of reinforcement for the S_D during testing would be expected to prolong the test process and thus increase the stability of the measures. By continuing the tests until discrimination was nearly perfect, it would be possible to determine whether or not the physical differences among the stimuli exceeded the sensory jnd. In addition, this operation would make it possible to examine the process by which the discrimination developed. Finally, the technique would help to maintain the behavior that had been developed during training. The maintenance of such control for the S_D seemed to be of special importance in assessing the gradients of approach and avoidance during conflict, for if either of the conflicting response tendencies with respect to the S_D was permitted to dissipate, or even change very much during the tests, it would greatly complicate the interpretation of the results.

Training on avoidance.—Ten new *S*s, maintained on a free feeding diet, were taught to press the bar to avoid electrical shock, the signal for which was the same 3500-cps tone which had been used to train approach. The average shock level was 100 v. ± 15 v. when measured across the bars of the grid floor. As in the experiments on approach, the tones appeared on an average of once every 35 sec. However, in this phase of the work, shock was programed to appear 4 sec. after the onset of tone and both the tone and the shock were programed to remain on until terminated by a bar press. A bar press during the 4-sec. warning period terminated the tone and prevented the occurrence of the associated shock. Thus, every tone was accompanied by either an avoidance response or an escape response.

To facilitate comparisons with the data on approach, it was desired that avoidance *S*s achieve a high level of performance prior to testing. For this reason, *S*s were required to meet a criterion of 90% of the shocks avoided on each of three consecutive sessions.

Only 5 *S*s were able to reach this criterion and they required an average of 24 sessions to do so (35 trials per session for the first 10 sessions and 70 trials per session for the rest). One *S* developed respiratory infection and was discarded. The other 4 *S*s failed to achieve the criterion even though they were exposed to 50 sessions of training. These *S*s were dropped from the present experiment but were employed in a sequence of studies designed to reveal the cause of their relatively poor performance. A detailed description of the development of avoidance for all 10 *S*s is presented in the paper which describes the work with the *S*s which failed to reach criterion (Hoffman, Fleshler, & Chorny, 1961).

Testing on avoidance.—The assessment of the gradient of avoidance involved the same tape recorded sequence of stimuli that was previously used to assess the gradient for approach. Except for the fact that a bar press during any tone always terminated that tone and that shock was programed to occur with all three 3500-cps tones in a given series, the testing procedure and the criterion for terminating the tests were identical to those used to assess the gradient of approach.

---

[2] At the start of this research, the investigators felt some concern as to whether the rat could actually hear pure tones with frequencies as low as 400 cps. Pilot work, however, revealed that they could: when rats were trained on approach using a 400-cps tone as the S_D and tested at higher frequencies, their gradients were approximately mirror images of those reported for approach in the present research.

Results

Gradient of approach and gradient of avoidance.—The mean probability of response for each of the several stimuli is presented for approach in Fig. 1 and for avoidance in Fig. 2, for selected sessions throughout testing. For these figures and for the associated analysis, the data for the 3500-cps tones were derived from the one S_D in each series which yielded no reinforcement for approach and from the 3500-cps tones which occupied identical serial positions in the tests of avoidance.

Because of the slow development of discrimination, the data from the first test session can provide a convenient index of the initial gradient of generalization. An analysis of variance of the response frequencies for the several stimuli from the first test session is presented in Table 1; it is based on a mixed two-factor design (Lindquist, 1953). The first factor, conditions, is a between-*S*s variable having two levels, approach and avoidance, (five rats in each group) while the second factor, stimuli, is a within-*S*s variable having six levels (the S_D and the five test tones).

Each of the *F*s in the analysis is highly significant. The main effect

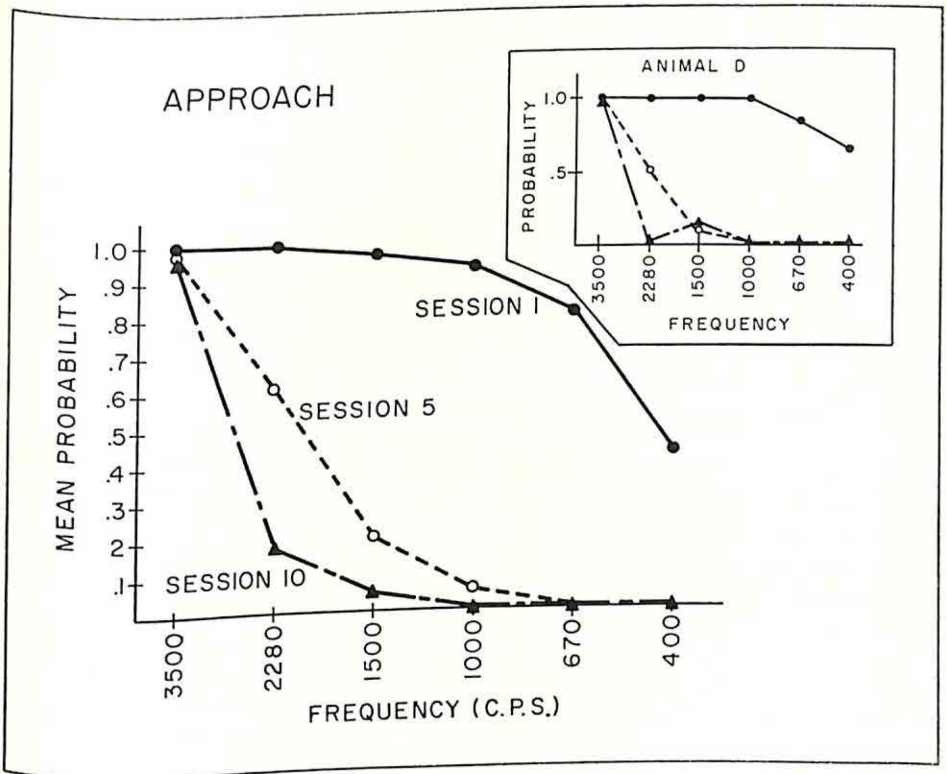


FIG. 1. The course of discrimination training for approach. (These functions show the mean probability of an approach response—plate press for food—to the S_D, 3500 cps, and to the five test frequencies during the first, fifth, and tenth session of testing. During the tests, food reinforcement was programed to occur with the S_D only. The inset shows the performance of a single animal from this group.)

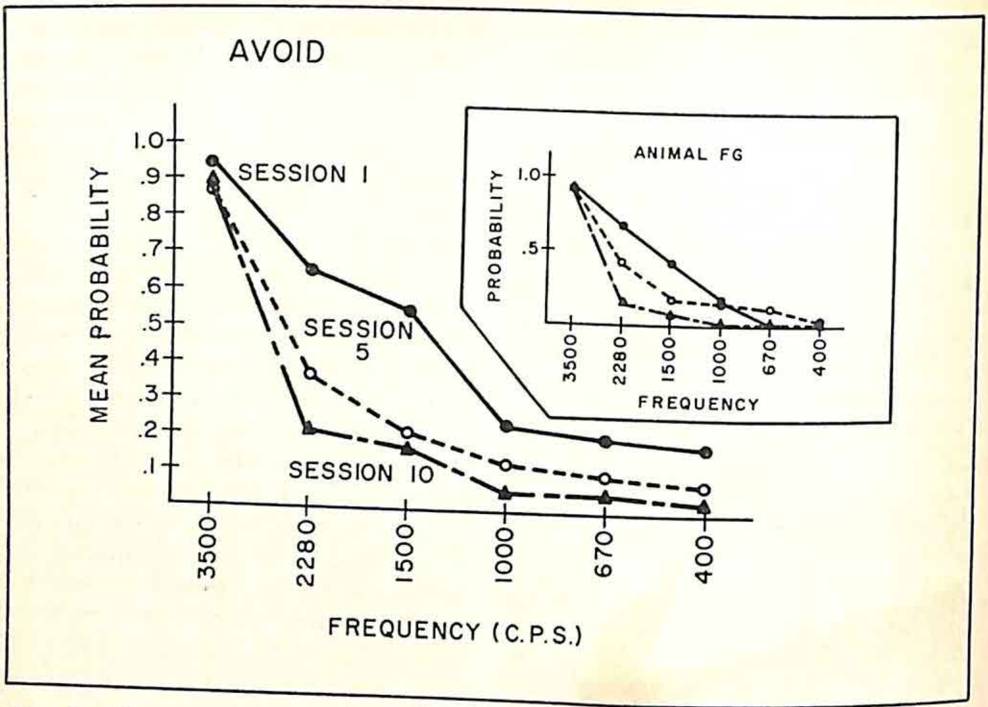


FIG. 2. The course of discrimination training for avoidance. (These functions show the mean probability of an avoidance response—bar press to prevent shock—to the S_D, 3500 cps, and to the five test frequencies during the first, fifth, and tenth session of testing. During the tests, shock was programed to occur with the S_D only. The inset shows the performance of a single animal from this group.)

of stimuli can be interpreted to result from the systematic generalization decrement which can be seen for both the approach and the avoidance gradient. The significant conditions effect reflects the overall higher level of the approach gradient. It can be seen in the figures that the avoidance gradient is sharper than the approach gradient. The significant level of the Stimuli × Conditions interaction indicates that, independent of the heights of the gradient, this observed difference in the shape of the gradient is significant. Apparently, in agreement with results obtained by Miller and his associates, approach generalized more broadly than avoidance.

In order to examine more carefully the course of responding to the nonreinforced test tones, a second analysis of variance was conducted on the response frequencies to stimuli other than the S_D for the first, fifth, and tenth test sessions (Table 2). Once

TABLE 1

ANALYSIS OF VARIANCE OF RESPONSE FREQUENCY: GENERALIZATION OF APPROACH VS. AVOIDANCE

Source	*df*	*MS*	*F*
Between *S*s	9		
Training conditions (C)	1	370.017	33.49***
Error (B)	8	11.467	
Within *S*s	50		
Stimuli (S)	5	79.310	26.96***
S × C	5	19.417	6.59***
Error (W)	40	2.947	
Total	60		

*** $p < .001$.

TABLE 2

ANALYSIS OF VARIANCE OF RESPONSE FREQUENCY: DEVELOPMENT OF DISCRIMINATION

Source	*df*	*MS*	*F*
Between *S*s	9		
Training conditions (C)	1	99.227	10.40*
Error (B)	8	9.543	
Within *S*s	140		
Stimulus (S)	4	92.343	33.40***
Test session (T)	2	511.286	71.20***
S × T	8	11.178	5.76***
S × C	4	4.443	1.69
T × C	2	165.486	23.10***
S × T × C	8	7.328	3.78**
Error (W)	112		
Error (W_s)	32	2.768	
Error (W_t)	16	7.178	
Error (W_{sxt})	64	1.940	
Total	149		

* $p < .05$.
** $p < .01$.
*** $p < .001$.

again, the analysis is a mixed factorial design with one between-*S*s factor, approach vs. avoidance, but in this case there are two within-*S*s factors, stimuli (the five test stimuli) and test sessions (1, 5, and 10). These data are also summarized in Fig. 1 and 2. The large number of highly significant main effects and interactions are indicative of a set of complexly changing relationships. The significant stimulus component reflects the substantial effect of the differences in acoustic frequency among the several test stimuli. Test sessions also have a large effect; as sessions progress, the overall frequency of generalized responses declines. The significant interaction between stimuli and test sessions indicates that the effect of sessions is at least partially determined by the acoustic properties of the tones.

The somewhat small main effect of conditions and the large interaction with test sessions yields to a straightforward interpretation (see Fig. 1 and 2); the significant, though small, main effect results entirely from the difference in the overall levels of approach and avoidance in Session 1; the large interaction is a result of the avoidance gradient being higher than that of approach in Sessions 5 and 10. In light of the previous analysis (Table 1), a somewhat surprising result is that the Stimulus × Conditions interaction only approaches significance. A careful examination of the figures reveals that although a large Stimulus × Conditions interaction exists in Session 1, in Session 5 it is small, and in Session 10 it is somewhat reversed. That this trend is significant is confirmed by the large triple interaction, which can be considered as the interaction of the Stimulus × Conditions interaction, with test sessions. Another way to consider the significant triple interaction is that the Session × Stimulus interaction differs for approach and avoidance.

The S_D responses were examined in a separate mixed factorial analysis of variance with conditions and sessions as the two factors. Only the training conditions' main effect (approach vs. avoidance) was significant, $F(1, 8) = 28.9$, $p < .01$. Therefore, it is clear that the test session effects revealed in the above analysis were not the result of changes in the stimulus control exercised by the S_D.

In reference to the generalized response tendencies, it may be concluded that although the effect of discrimination training (test sessions) was to reduce the response tendency to stimuli on the wing of the generalization gradient, its effects were not evenly distributed among these stimuli nor were they the same for approach and avoidance; with discrimination training the gradient of approach sharpened more rapidly than the gradient of avoidance and this effect

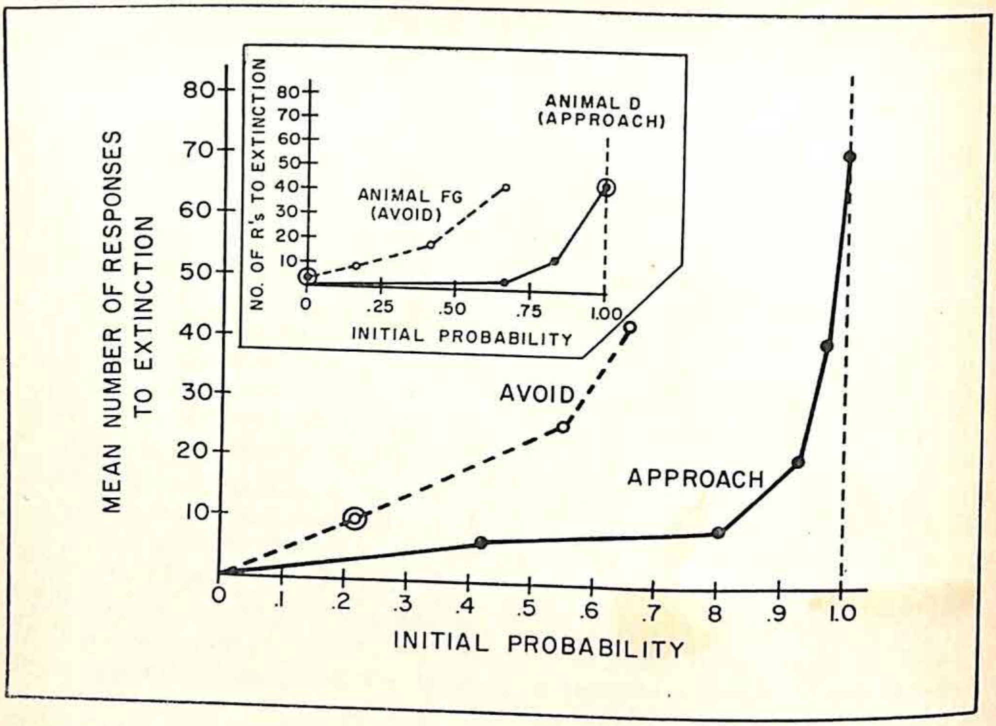


FIG. 3. The mean number of responses to extinction plotted as a function of the initial probability of response. (The coordinates for each point represent the probability of response to a given stimulus other than the S_D during the first session of testing—initial probability—and the total number of responses to that stimulus throughout the course of discrimination training—responses to extinction. For example, during the tests on approach the 670-cps tone produced an initial probability of .8 and, on the average, yielded nine responses throughout discrimination training. The circled points represent averages for those stimuli which yielded similar initial probabilities. The inset shows these functions for the individual animals that were represented in the insets of Fig. 1 and 2.)

is evident despite the finding that, initially, approach generalized much more broadly than avoidance.

Additional data bearing on this question are provided by the finding that for approach *S*s, the discrimination criterion was met after an average of 11 sessions, whereas avoidance *S*s required an average of 14 sessions to meet the same criterion. Figure 3 permits a more careful analysis of the discrimination process. It shows for each stimulus on the wing of the gradient, the total number of generalized responses to the discrimination criterion (extinction) when plotted as a function of the response probability observed in the first session for that stimulus. For both conditions the total number of unreinforced responses increased as a function of the initial probability; however, at a given initial probability, resistance to extinction was greater for avoidance than for approach. This relationship, which is not readily seen in the other figures, is clearly representative of large differences existing in the processes which underlie the development of discrimination in approach and in avoidance.

EXPERIMENT II: ASSESSMENT OF SIMULTANEOUS GRADIENTS OF APPROACH AND AVOIDANCE DURING APPROACH-AVOIDANCE CONFLICT

This phase of the research involved five steps and required that the *S*s meet a sequence of rigid behavioral criteria. The establishment of the criteria was necessary in order to make certain that, prior to the establishment of conflict, the performance of the *S*s would be comparable to the performance of animals tested on approach only and on avoidance only. Since the previous work had indicated that one could not accurately predict which *S*s meet the criterion for avoidance, it was decided to develop conflict by training avoidance first, and thus minimize the chances of investing large amounts of time in *S*s which could not meet the criterion. Out of an initial pool of 10 *S*s, only 3 completed all phases of the conflict training. Four of the *S*s failed to meet the criterion for avoidance, and 3 of the *S*s were lost because of respiratory infections.

Method

Development of avoidance behavior.—The procedures described previously were used to establish a performance in which *S*s avoided shock by pressing the bar upon the presentation of a 3500-cps tone. As with *S*s tested only on avoidance, the criterion performance was three consecutive sessions in which the *S* avoided at least 90% of the shocks. For the 3 *S*s which completed the training sequence, this initial condition was met within 22 sessions and the acquisition function, as well as the asymptotic behavior of these *S*s, was essentially the same as that of the *S*s that had been used to assess the gradient of avoidance.

Extinction of avoidance behavior.—The tone that had been used in training avoidance was presented without shock until the *S*s failed to respond throughout an entire session; this procedure required three sessions.

Development of approach behavior.—The *S*s were gradually deprived of food until their body weights stabilized at 80% of their previously determined free feeding body weights. During the period of weight reduction, the *S*s were placed in the experimental chamber for 2 hr. per day, to permit generalized emotional responses to extinguish. They were then trained to press the plate in the presence of the same 3500-cps tone that had previously been used as a warning for impending shock. The training procedure and the criterion for the termination of training were identical to those previously employed in the development of approach only. This phase of the training required an average of 16 sessions and yielded a performance which was essentially the same as that of *S*s trained on approach only.

Development of conflict.—To establish conflict, shock was again programed to begin 4 sec. after the onset of the tone. The stimulus-response contingencies were such that an initial bar press during tone, but prior to the onset of shock, terminated that tone and prevented the occurrence of the shock. An initial plate press during tone, on the other hand, provided immediate food reinforcement (40% of the time) but it did not terminate the tone nor did it prevent the associated shock. In addition, if during the 4-sec. warning period a plate press occurred and was followed by a bar press, the plate press might produce immediate reinforcement (40% of the time) but the bar press would not terminate the tone nor would it prevent the associated shock. Finally, whenever shock came on, both it and the tone remained on until a bar press occurred. Thus, for each tone, the *S* was placed in conflict. An initial plate press might produce food but it would prevent the *S* from avoiding the associated shock. An initial bar press, on the other hand, would prevent shock, but it would terminate the tone and, hence, remove the opportunity to obtain food. As in other phases of the work, 70 tones occurred on each session and the interval between tones was variable (average interval equal to 35 sec.).

With the reintroduction of shock at the end of tone, avoidance was almost immediately reinstated so that only two sessions of conflict training were conducted. During the second session, a performance was exhibited in which (totaling across all 3 *S*s) 86% of the tones yielded an initial avoidance response (bar press) whereas only 7% of the tones yielded an initial approach response (plate press) and of these, only 12% were followed by a bar press prior to the onset of shock.

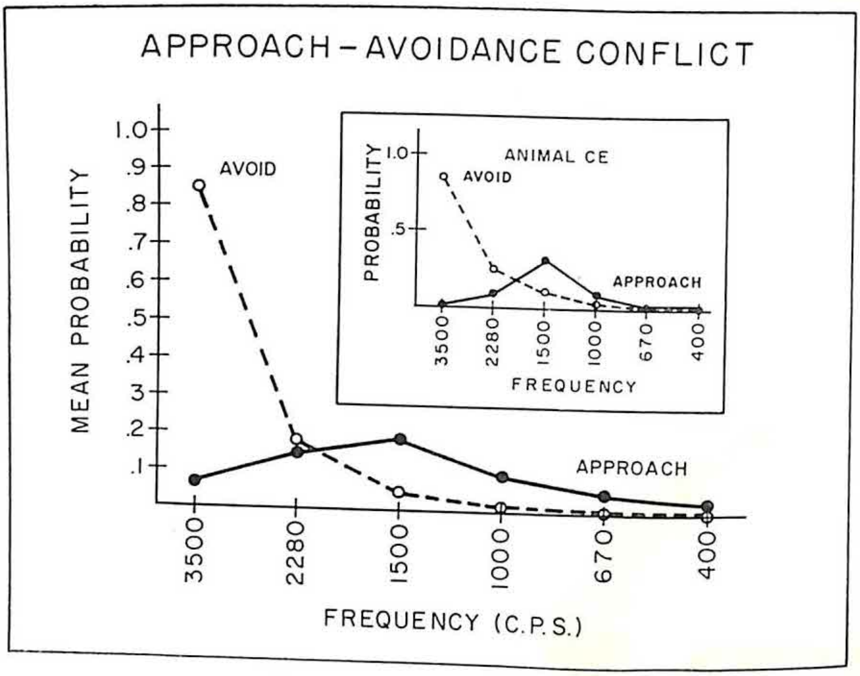


FIG. 4. The generalization gradients of approach and of avoidance, during conflict. (These functions show the mean probability of approach and avoidance responses—plate press for food and bar press to prevent shock—to the S_D, 3500 cps, and to the five test frequencies during the initial session of the tests for stimulus generalization under conflict. During this session, food reinforcement and shock were programed to occur with the S_D only. The inset shows the performance of a single animal from the group.)

Test for stimulus generalization under conflict.—The procedure was similar to that used in the assessment of the generalization of approach and of avoidance when measured separately and involved the same sequence of tape-recorded stimuli. In a given series, shock was programed to occur with all three 3500-cps tones, and food reinforcement was programed for two of these. In all other respects, the stimulus-response contingencies with reference to the 3500-cps tones were identical to those which were in force during the development of conflict. The other tones in the series were programed to remain on for 5 sec. (unless terminated by a bar press which had not been preceded by a plate press) and neither shock nor food reinforcement occurred in their presence. As in previous tests, the entire series of eight tones was presented 12 times per session.

Results

Although 10 test sessions were conducted, only the results of the first session exhibited systematic generalization effects. After Session 1, very few generalized responses occurred, either for approach or avoidance, and such responses as did appear seemed to be distributed randomly among the several test stimuli. At present, the reason for this effect is unclear.

The results for the first session of testing are depicted in Fig. 4. The avoidance response occurred with a high probability to the S_D and few avoidance responses occurred to the other test stimuli. The mode of the avoidance gradient was the S_D (3500 cps) whereas the mode of the approach gradient was displaced to the 1500-cps tone.

To assess the reliability of the observed displacement effect, a *t* test for related observations was conducted to compare the sum of the generalized approach responses to the sum of the generalized avoidance responses during conflict. This test yielded a $t(2) = 4.401$, $p < .05$ for a two-tailed test. A second t test for related observations was conducted on the approach vs. avoidance responses to the S_D. This test yielded a $t(2) = 21.0$, $p < .01$ for a two-tailed test.

Apparently the effects seen in Fig. 4 are reliable, despite the small number of *S*s in the sample. When the responses were placed in conflict a behavior emerged in which the dominant response to the S_D was avoidance, whereas the dominant response to stimuli on the wing of the generalization gradient was approach.

To assess the effects of the approach-avoidance conflict on avoidance behavior, the avoidance gradient in Fig. 4 was compared to the gradient during the first session of testing for *S*s trained on avoidance only (Fig. 2).

It appears that the conflict between approach and avoidance had little, if any, effect on avoidance responses to the S_D (3500 cps), $t(6) = 1.14$ ($p > .05$) for a two-tailed test. For stimuli on the wing of the gradient, however, the conflict between responses produced a condition in which fewer generalized avoidance responses occurred than during the comparable test for avoidance only. To assess the reliability of this difference, a t test was conducted on the total number of the generalized avoidance responses during the first session of testing in the absence of conflict as opposed to the comparable data obtained during conflict (Fig. 4). The results of this comparison were statistically significant, but barely so, $t(6) = 2.448$, $p < .05$ for a two-tailed test.

TABLE 3

LATENCIES OF APPROACH AND AVOIDANCE UNDER INDEPENDENT AND CONFLICT CONDITIONS

Conditions	Mean (Sec.)	*SD* (Sec.)	*N*
Conflict			
Approach	1.38	.234	3
Avoidance	2.52	.283	3
Independent			
Approach	1.46	.292	5
Avoidance	2.49	.32	5

As can be seen in the comparison between the first session data in Fig. 1 and the data in Fig. 4, the dominant effect of conflict was to attenuate the tendency to make approach responses. A t test on the frequency of approach responses to the S_D during, and in the absence of, conflict yielded a $t(6) = 11.42$ ($p < .01$ for a two-tailed test). Similarly, a t test on the sums of the generalized approach responses during, and in the absence of, conflict yielded a $t(6) = 11.02$, $p < .01$ for a two-tailed test.

Clearly, the presence of a competing approach response had only a small effect on the gradient of avoidance, whereas the presence of competing avoidance responses had a large effect on the gradient of approach. This finding is of special interest, in view of the fact that when tested separately, the initial gradient of approach was everywhere higher than the initial gradient of avoidance.

Latency measures were obtained from the second conflict session (the last session prior to testing). A t test on the difference between the mean latency of approach and that of avoidance was computed for each *S* and was found to be significant in each case ($p < .01$). It was also found, however, that the mean latencies for approach and avoidance under conflict were very similar to those that were obtained for approach *S*s and for avoidance *S*s when independently measured during their last session prior to testing. These data are shown in Table 3. All possible *t* tests were conducted; only those involving an approach vs. an avoidance group were significant (each of these attaining at least the .05 level for a two-tailed test). Thus, while it can be concluded that the avoidance response has a longer latency than the approach response, it appears that conflict between the responses had essentially no effect on the latency characteristics of either response.

DISCUSSION

In order to facilitate comparisons between the groups (approach vs. avoidance vs. conflict) only *S*s which exhibited roughly comparable behaviors were tested. As a result, while the several groups are in one sense equated, they also differ in certain respects. For example, the rate of attrition during training was much higher for avoidance than for approach. Also, the conditions of training were not identical for the two responses. Not only were different patterns of reinforcement used for approach and avoidance but the manipulandum for the two behaviors was not the same. The manipulanda proper are probably unimportant since pilot work on approach indicated that there were essentially no differences in rate of acquisition, asymptotic probability, or latency as a function of bar press vs. plate press. The differences in pattern of reinforcement, however, pose a problem which, unfortunately, cannot be resolved at this time. In order to match the treatments with respect to pattern of reinforcement, it would be necessary to identify precisely the reinforcing event in avoidance and, at present, knowledge of this event is incomplete.

Because of ambiguities such as these, a certain degree of caution must be exercised when interpreting the results of this work. Even so, however, the broad implication of the results is perfectly clear. When *S*s are trained to perform approximately the same on avoidance as on approach, subsequent tests for stimulus generalization and discrimination uncover large differences in behavior.

In certain respects, the combined results of Exp. I and II are in close agreement with predictions based upon Miller's theory. For example, the theory holds that if displacement is to occur, the gradient of avoidance must be steeper than the gradient of approach. In the present study, displacement was observed and the gradient of avoidance was, in fact, steeper than the gradient of approach. Certain aspects of the results, however, disagree with the assumptions which underlie the theory. For example, according to the theory, displacement will occur only when the gradients cross. In the present study, displacement occurred despite the finding that when tested separately the gradient of approach was everywhere higher than the gradient of avoidance. Finally, the theory implicitly assumes that when approach and avoidance are placed in conflict, each will inhibit the other, according to the relative strengths of the two responses. In the present study, however, when measured separately, approach was stronger than avoidance. Yet, when placed in conflict, approach showed a large decrement, whereas avoidance was only slightly affected.

Previous research (Hoffman & Fleshler, 1962) is relevant to the issue. In that study, avoidance was developed while *S*s were engaged in positively reinforced ongoing behavior. It was found that the tendency to avoid was related to, and indeed, seemed to depend upon, the capacity of the warning signal to suppress the ongoing behavior. It is reasonable to suppose that during the conflict phase of the present study, the large decrement in approach also represented a conditioned suppression phenomenon, and hence, that any direct competition between approach and avoidance tendencies per se was probably of only minor importance. In part, this interpretation is supported by the additional finding that when measured separately, the latency of avoidance was consistently larger than the latency of approach and that neither latency was modified during conflict.

The findings with reference to latency are of interest in a second respect. Basically, they mean that during conflict, where either approach or avoidance might occur, the probability of a resolution in favor of approach was highest at the beginning of the conflict situation. On the other hand, while the probability of a resolution in favor of avoidance was low at the beginning of the conflict situation, this probability increased as conflict persisted. In essence, during approach-avoidance conflict, approach tended to occur early in the conflict situation or not at all, whereas avoidance tended to occur late or not at all.

REFERENCES

HOFFMAN, H. S., & FLESHLER, M. The course of emotionality in the development of avoidance. *J. exp. Psychol.*, 1962, **64**, 288–294.

HOFFMAN, H. S., FLESHLER, M., & CHORNY, H. Discriminated bar-press avoidance. *J. exp. Anal. Behav.*, 1961, **4**, 309–316.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

(Received April 27, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 292–296

HUMAN REINFORCEMENT: VICARIOUS AND DIRECT[1]

FREDERICK H. KANFER AND ALBERT R. MARSTON
University of Oregon Medical School *University of Wisconsin*

4 experimental groups of 9 Ss each said words intermittently with a taped simulated group in a verbal conditioning task. In a factorial design groups were given vicarious reinforcement (VR), direct reinforcement (DR), both, or neither. VR consisted of Ss listening to E's "Good" when a taped S gave a critical response. DR consisted of saying "Good" directly to S. Control groups heard tapes with fewer critical responses and no VR, or no tape. Half of the Ss in each condition were directly reinforced, half were not reinforced. In extinction no tape was played. VR significantly facilitated learning in the experimental groups, with DR showing no additional effects. The control groups failed to learn. Extinction differences were due to differences in acquisition levels.

One source of complexity in human learning lies in *S*'s ability to interrelate informational inputs, to make covert observations of contingencies between his own behavior and external events, and to assess the likelihood of success prior to making an overt response. Of special interest to us have been the effects of information, both by instruction and by structure of the task (Kanfer & Marston, 1961, 1962), on verbal learning. In the absence of direct reinforcement of overt behavior an *S* may still be able to learn either by self-reinforcement (Kanfer, Bradley, & Marston, 1962; Skinner, 1953), or by observation of response reinforcement contingencies in others, as in vicarious reinforcement (Berger, 1961; Lewis & Duncan, 1958; Miller, 1961). In both instances information about the appropriateness of the response is conveyed *not* by the usual reinforcing procedure. These instances deviate from the typical procedure in that either *E* administers reinforcement to another person but not to *S* (vicarious reinforcement); or *E* relinquishes control over the administration of reinforcement to *S* (self-reinforcement). It is the first instance which was examined in the present study.

The purpose of this experiment was to study the effects of vicarious reinforcement on verbal learning in simulated groups, in order to ascertain whether observation of reinforcement for learning in others is an effective reinforcing event for human *S*s.

METHOD

Subjects.—Eighty-five students from an introductory psychology course served as *S*s. Half of the *S*s were males, half were females.

Apparatus.—The *S*s were seated at a table in a darkened room. On a 2 × 4 ft. black panel in front of them was mounted a small signal light. To the left of the light a microphone rested in a slot on the board. The *S* wore earphones throughout the experiment. The earphones and microphone were part of an intercommunication system with an adjacent room in which *E* monitored the experiment. A Wollensak (Model T-1500) tape recorder served as an amplifier in the system. With his own microphone and earphones, *E* was able to hear *S*, to speak to him, and to play the tape recorded "group" responses into both earphones. The *E* used a telegraph key to activate the light on *S*'s

---

[1] This study was supported in part by Research Grant MH-06922-01 from the National Institute of Mental Health, United States Public Health Service. The authors wish to acknowledge the participation of D. S. Leventhal and D. R. Slavin who assisted in this study.

TABLE 1

EXPERIMENTAL DESIGN

Group	Condition		N
	Tape	Rf to S	
Experimental			
A	Acquisition, VR	DR	9
A′	Acquisition, VR	—	9
B	Acquisition, No VR	DR	9
B′	Acquisition, No VR	—	9
Control			
C	Operant, No VR	DR	9
C′	Operant, No VR	—	9
D	No tape	DR	9
D′	No tape	—	9

board. The tapes were produced by 9 graduate students who re-enacted a group conditioning session similar to that described by McBrearty, Marston, and Kanfer (1961) in which human responses were reinforced. Words from various parts of actual group protocols were read in turn by each student.

Design and stimuli.—The experiment consisted of a 2 × 2 factorial design and four control groups (see Table 1). The tape content determined the presence or absence of vicarious reinforcement (VR). The same tapes were used for Groups A and A′; B and B′; C and C′. For the A groups the tape contained words taken from an acquisition record of an earlier study. Fifty percent of the responses were human nouns and were followed by *E* saying "Good." As in actual acquisition sessions, the proportion of critical responses increased over blocks of trials. For the B groups the same words were read but the tape contained no reinforcing stimuli. For the C groups a tape was prepared on which the words were taken from a preacquisition (operant) phase of a previous study. Only 5% of the words were human nouns. For the D groups no tapes were played. The experimental groups yielded a comparison of the effects of vicarious reinforcement on a series of 30 acquisition and 30 extinction trials, both with and without the addition of direct reinforcement (DR). Groups C and C′ provided controls for the tape content and D and D′ for the administration of any recorded stimuli.

Procedure.—When *S* arrived in the hall he found a row of office doors with "Experiment in Progress" signs. He was ushered into the experimental room by *E* and asked to wait a few moments until the (nonexistent) "other *S*s" had been instructed. When *E* returned, he told *S* that this study investigated how people use words. The *S* was asked to say words, any words except proper names, phrases, or numbers. The *E* told *S* that there were a number of *S*s in similar cubicles in other rooms performing the same task. The *S* was asked to wear earphones in order to "hear some of what was going on in the other rooms." The *E* continued: "When it is your turn, your light will blink. This is your signal to say a word. Give the first word you think of." The same instructions were given to all eight groups.

All *S*s who in a postexperiment questionnaire gave any indication that they might have suspected hearing a tape recording rather than live voices were eliminated from the study. Thirteen *S*s were discarded for this reason, apparatus failure, or other reasons, leaving 72 *S*s. The *E* began conditioning by activating the signal light 10 times at approximately 12-sec. intervals. These 10 trials constituted a measure of the operant frequency of human nouns.

Following the operant trials, except in Groups D and D′, *E* turned on the recorder and *S* heard the first round of nine responses of his "group." At the end of the ninth taped response, or after the subsequent "Good" if appropriate, the "Stop" button was pushed, and *S* was signaled to respond. This sequence continued for 30 rounds of trials. For Groups

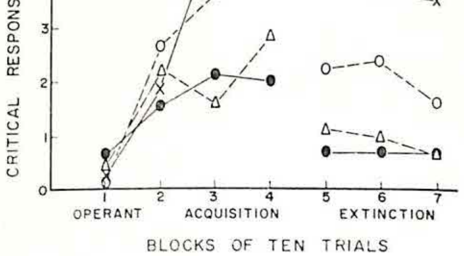


FIG. 1. Acquisition and extinction curves for the experimental groups. (All groups heard acquisition tapes under the following conditions: A-VR and DR; A′-VR only; B-DR only; B′ neither VR nor DR.)

TABLE 2

Analyses of Variance on Acquisition

Source	Acquisition			Extinction		
	df	*MS*	*F*	*df*	*MS*	*F*
VR vs. No VR	1	52.56	4.30*	1	131.12	7.00*
DR vs. No DR	1	2.50	<1.00	1	30.08	1.60
VR vs. DR	1	0.35	<1.00	1	14.09	<1.00
*S*s in groups (*S*s)	32	12.20		32	18.73	
Blocks (B)	3	80.86	23.78**	2	1.73	<1.00
B × VR	3	16.90	4.97**	2	0.62	<1.00
B × DR	3	1.92	<1.00	2	0.86	<1.00
B × VR × DR	3	6.04	1.77	2	0.25	<1.00
B × *S*s	96	3.40		64	1.59	

Note.—Including one operant block and extinction.
* $p < .05$.
** $p < .01$.

D and D′, *E* continued to activate the signal light at 12-sec. intervals, approximating the time required for one round of the taped voices. Group D heard only *E*'s reinforcement when appropriate, and D′ heard nothing over the earphones.

Following acquisition all *S*s received 30 extinction trials during which *E* activated the light at 12-sec. intervals without reinforcement. At the end of the experiment, *E* administered a questionnaire which asked *S*: (*a*) What he thought the purpose of the experiment was; (*b*) Whether it was easy or difficult and why; (*c*) How he chose his words; (*d*) What he heard through the earphones; (*e*) How what he heard affected his behavior; (*f*) How many other people he heard; (*g*) Whether he thought that he did better or worse than the others; and (*h*) How the voices sounded to him. The last three questions were omitted for Groups D and D′.

RESULTS

Acquisition.—The frequency of correct responses was analyzed in blocks of 10 trials. The learning curves for the four experimental groups are shown in Fig. 1. The curves show a similar performance in all groups during the operant period with subsequent differentiation among the groups. A summary of the analysis of variance on the acquisition and extinction scores is presented in Table 2. The addition of vicarious reinforcement on the tapes (Groups A and A′) resulted in significantly more learning ($F = 4.30$) and significantly greater learning increments over blocks ($F = 4.97$), thus demonstrating the facilitating effects of VR in this task. Combining all groups, the significant Blocks effect suggests that performance improved with increasing number of trials. An $F < 1.00$ for the DR variables indicates that the addition of direct reinforcement does not significantly enhance performance, when compared to Groups A′ and B′ which did not receive DR.

To establish the minimal conditions of reinforcement and tape content necessary for statistically significant evidence of learning in this task, Blocks × *S*s analyses were run for all eight groups over operant and acquisition blocks. With 3 and 24 *df*, the following *F* values were obtained for Groups A, A′, B, and B′ in their respective order: 12.32 ($p < .001$); 8.21 ($p < .001$); 4.02 ($p < .025$); 1.61 ($p > .20$). None of the *F* values in the four control groups approached significance at $p = .20$. These results indicate that only those groups which were exposed to acquisition tapes and received either DR or VR showed learning. In considering the failure of DR to produce learning in groups exposed to tapes of operant sessions or no tapes at all it should be remembered that in the usual verbal conditioning procedure *S*s emit responses at the rate of approximately 20 words/min for many minutes before learning is measured (Greenspoon, 1955; Kanfer, 1958). In this study *S*s had only 30 opportunities for reinforcement of their responses in acquisition. Thus the failure to find learning in the control groups under these conditions merely emphasizes the great accelerating effects of VR.

Although neither Groups B′ nor C′ showed large enough increases over blocks to yield evidence of learning, an *F* test between these groups was carried out to determine the effect of tape content. An *F* of 19.61 ($p < .001$) indicates that exposure to a tape containing about 50% critical responses resulted in emission of significantly more critical responses than exposure to a tape with 5% critical responses. The nonsignificant Blocks × Groups interaction in this analysis suggests no greater increment but only a higher overall level of responding in Group B′.

Extinction.—Table 2 presents an analysis of variance on the extinction data. As in acquisition, the VR groups yielded more responses than No VR groups. It can be seen from Fig. 1 that the order of the experimental groups is the same as in acquisition. These group differences may have been due to differential performance in acquisition rather than the effect of the extinction per se. An analysis of covariance on extinction scores was performed, adjusting for the total number of correct responses in acquisition. The adjusted *F* values for the VR and DR variables and for their interaction were not significant at $p = .05$, indicating that the differences among VR and No VR groups in extinction could be attributed mainly to the different performance levels of these groups in acquisition.

Because of the similarity of the present procedure to that used in earlier studies of verbal conditioning, the relationship of awareness to learning was also examined. The replies to the postexperimental questionnaire were rated as to degree of awareness on a scale developed in previous studies (Kanfer & Marston, 1961). On a six-point scale, a rating of 1 indicated no reference to the reinforcement nor to the correct class, a 2 rating indicated reference to the reinforcing stimuli. A 3 rating required reference to the correct class. *t* tests were computed for each group to test deviation from a mean rating of 1. Only Group A yielded a significant *t* (3.50, $p < .01$), with a mean of 3.78. Group A′ was the only other group whose mean (2.01) exceeded a 2 rating.

DISCUSSION

Exposure to contingencies between a critical response class and a social reinforcer in others served as an effective reinforcing stimulus for the observing *S*. The question arises whether VR contributes to learning mainly because of its informational aspects rather than its reward value. The variation in tape content in this study yields some evidence on the effects of making available to *S* high (50% in the experimental groups) or low (5% or 0% in the controls) degrees of occurrence of critical responses as environmental cues. When neither VR nor DR is given (Group B′) it is seen that the informational input alone does increase *S*'s use of critical responses. However, this increase falls short of statistical significance. The *S*s in Group B′ reach a maximum frequency of about 20% correct responses when listening to the tape containing critical words in proportions of 36%, 51%, and 62%, respectively, over three acquisition blocks. Thus, in the absence of both VR and DR, *S*s failed to match the model in the manner described as imitation by Miller and Dollard (1941). On the other hand, Groups A and A′ approximated the mean frequency of critical responses on the tape during each trial block fairly well (see Fig. 1). Nonreinforced imitation alone, at least in the present task, was not a sufficient condition for acquisition.

The large initial drop from the end of acquisition to extinction probably reflects in part *S*'s discrimination of the procedural change. This transition was clearly signaled by the cessation of input from the simulated group. The analysis of the extinction data indicates however that some of the training effects persisted. Further, as shown in Fig. 1, the groups do not return to the preacquisition level, at least for the time sampled in this extinction procedure.

It was noted that direct reinforcement did not result in learning when an "operant" tape or no tape was played. Thus, the use of VR and an "acquisition" tape considerably lowers the number of trials required for learning. However, it should be considered also that only few *direct* reinforcements could be obtained for the 30 responses made by *S*, in contrast to the 270 taped responses.

If a mediational framework is adopted (Berger, 1961; Lewis & Duncan, 1958), it may be assumed that *S* rehearses or responds covertly as he listens to others. If he tells himself "I was right" or in some other way, verbal or not, administers some reinforcement to himself, vicarious training may directly contribute toward increased response strength in *S* during vicarious trials. That information by self-reinforcement may have some effects on response strength is shown in a study by Marston and Kanfer (1963) which further explores the process of self-reinforcement.

REFERENCES

BERGER, S. M. Incidental learning through vicarious reinforcement. *Psychol. Rep.*, 1961, **9**, 477–491.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

KANFER, F. H. Verbal conditioning: Reinforcement schedules and experimenter influence. *Psychol. Rep.*, 1958, **4**, 443–452.

KANFER, F. H., BRADLEY, M. M., & MARSTON, A. R. Self reinforcement as a function of degree of learning. *Psychol. Rep.*, 1962, **10**, 885–886.

KANFER, F. H., & MARSTON, A. R. Verbal conditioning, ambiguity, and psychotherapy. *Psychol. Rep.*, 1961, **9**, 461–475.

KANFER, F. H., & MARSTON, A. R. The effect of task-relevant information on verbal conditioning. *J. Psychol.*, 1962, **53**, 29–36.

LEWIS, D. J., & DUNCAN, C. P. Vicarious experience and partial reinforcement. *J. abnorm. soc. Psychol.*, 1958, **57**, 321–326.

MCBREARTY, J. F., MARSTON, A. R., & KANFER, F. H. Vicarious experience in a group conditioning situation. Paper read at American Psychological Association meetings, New York, 1961.

MARSTON, A. R., & KANFER, F. H. Human reinforcement: Experimenter and subject controlled. *J. exp. Psychol.*, 1963, in press.

MILLER, N. Acquisition of avoidance dispositions by social learning. *J. abnorm. soc. Psychol.*, 1961, **63**, 12–19.

MILLER, N., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1953.

(Received April 27, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 297–304

VALUE OF STANDARD AND VERY FIRST VARIABLE IN JUDGMENTS OF REFLECTANCE OF GRAYS WITH VARIOUS RANGES OF AVAILABLE NUMBERS

E. C. POULTON AND D. C. V. SIMMONDS[1]

Applied Psychology Research Unit, Cambridge, England

1200 unsophisticated O's each made 2 numerical estimates (NEs) of the magnitude of the reflectance of a variable gray with respect to a standard typically called 10. With standards of high reflectance, variables with reflectances still higher gave the larger exponents. The reverse was the case with standards of low reflectance. The effects increased as the range of numbers available to O was reduced. Averaging exponents from multiple and fractional estimates produced larger means when the reflectance of the standard was high. Fractional NEs were influenced by prior multiple NEs and vice versa. These results raise doubts about the generality of direct measurements of sensory magnitude.

Over the last few years a considerable edifice of psychophysical functions for sensory magnitudes has been erected using the method of numerical estimation popularized by S. S. Stevens. In recent publications Stevens (1960a, 1960b, 1961a, 1961b) has listed values of some 22 exponents for various sensory continua, to which the power of the physical magnitude of the stimulus has to be raised in order to match a constant multiple of the magnitude estimates made by groups of *O*'s. He has also stated his belief that these exponents may reflect differences in the nature of the various sensory "transducers" or end organs used in making the estimates of sensory magnitude (Stevens, 1960b, p. 251; 1961a, p. 27).

Prior context.—While the construction work has been under way, cracks have appeared from time to time in the foundations. These have been surveyed, and the arguments put forward that they are not serious threats to the superstructure. The first crack appeared as a result of Garner's (1954, Table 1; 1959, Fig. 2) work on the effect of prior context in judgments of more than or less than half as loud. Garner showed that by selecting different ranges of variables to present to different groups of naive *O*'s, they could be made to report half loudness when on an average the ratio of the standard intensity to the variable intensity was approximately 10, 20, or 30 db.

It can be argued that this was because Garner used the constant method, in which *O* was allowed to state only whether the variable was more or less than half as loud as the standard. Being somewhat unsure at the start as to the exact value of half loudness, *O* was not very likely to use exclusively one of the two possible categories of judgment. Thus the average judgment of half was almost bound to fall somewhere within the range of stimuli presented, however unlikely the range was.

---

[1] We are grateful to S. S. Stevens for a letter suggesting the comparison which led to the present series of experiments, to the authorities and student volunteers of the Cambridge colleges in which the experiments were conducted, to P. Bakan for commenting on a draft manuscript, and to the British Medical Research Council for financial support.

It can also be argued that the overruling effect of prior context demonstrated by Garner should not occur with numerical estimates (NEs) of magnitude, since *O* is free to choose any number within the range available to him. However recent evidence suggests that NEs are not completely unaffected by prior context. Warren and Poulton (1960) and Poulton and Webster (unpublished)[2] using unpracticed *O*'s found that the average NE of the reflectance of a second gray variable was often significantly influenced by the reflectance of the variable judged first. Stevens and Stevens (1960, pp. IX-8–12), Künnapas (1960), and Strangert (1961) have all reported that the NEs of sophisticated *O*'s can be influenced by the range of variables presented to them.

Clearly the only method of eliminating the possible effect of prior context is to use the very first judgments of unsophisticated *O*'s, which means having a different group of naive *O*'s for each value of the variable. This was the answer which Stevens and Poulton (1956) formulated to Garner's criticism of all direct methods of judging sensory magnitudes, since *E* now no longer influences *O* by his choice of the range of stimuli. However it is not a complete answer, since *E* has still to select the value of the standard.

Choice of standard.—Stevens (1955) has collected together data showing the ratio in db. between a standard and a variable sound which has been found to be required for half and twice loudness at various sound pressure levels of the standard. He showed that with loud standards only a relatively small increase in intensity is required for twice loudness, whereas a relatively large decrease is required for half loudness. The reverse is the case for soft standards, Where *O* adjusted a variable sound to half or twice the loudness of the standard, Stevens discussed the effect in terms of the avoidance of uncomfortably loud and uncomfortably soft sounds. Where *O* estimated numerically the loudness of a variable presented to him, the effect can be discussed in terms of a confusion between the absolute magnitude of the variable and its magnitude relative to the standard, the NE being a compromise between the two (Stevens, 1956, pp. 6–8).

Since with sophisticated *O*'s the effects are not very large, and the effect for fractional judgments is in the reverse direction to the effect for multiple judgments, Stevens suggested that the effects should more or less balance each other out when the two sets of data are combined. The experiments to be reported show that the effect is considerably more marked with the very first NEs of untrained *O*'s who are not constrained by prior context, and that the effects of multiple and fractioned NEs are not always equal and opposite.

METHOD

Observers.—These were mainly male undergraduates of Cambridge University, although a few male graduates and women students were included. None had previously judged sensory magnitudes, and each *O* served in only one experimental condition. Students in psychology were not used.

Display.—A brown table top 27 × 27 in. with a reflectance of 20% always served as a background. In the middle of it were placed a standard and a variable patch each 11 × 6 in. The patches lay together with a long side contiguous, and short sides closest to *O*. Typically the standard patch was labeled by a 1-in. square of the same reflectance, which carried the number 10 written in pencil and lay immediately below and touching the center of the short side nearest to *O*. Black standards were labeled by a square of brown material similar to the table top, again carrying the number in pencil. With four extra groups of *O*'s the number 10 was changed to 1 or 1,000. For the first half of the *O*'s in each group the standard lay on the left of the variable, and for the second half it lay on the right.

---

[2] Manuscript entitled "Value of First Few Variables and Background in Fractional Estimates of the Reflectance of Grays."

The white and gray patches consisted of paper coated with nonglossy paint, supplied by the Mitchell Colour Card Company, London. Black cotton and rayon velvet were also used. The reflectances were measured both with a Macbeth illuminometer, and with an illuminometer made by Salford Electrical Instruments Limited. The lighting on the table was at the ambient level to which *O* had become adapted. (This appeared preferable to introducing a standard extraneous source of light. A previous experiment using separate groups of 50 *O*'s had shown that judgments of relative reflectance made in bright daylight were no different from judgments made in very much less bright artificial light.)

Experimental conditions.—These are listed in Tables 1 and 2. Half the groups estimated relative lightness and half relative darkness. This altered the range of numbers available, since a fractional estimate of lightness corresponds to a multiple estimate of darkness, and vice versa (Torgerson, 1960). Each *O* made two NEs, typically one fractional estimate and one multiple estimate, but when the standard was at either extreme end of the range both estimates had to be fractional or multiple. The order of presentation of the variables was counterbalanced for different groups of *O*'s. Thus Groups IIB and IID estimating relative lightness counterbalanced each other, and so did the comparable Groups IIA and IIC estimating relative darkness. Table 3 shows the four additional groups of *O*'s for whom the standard was called 1 or 1,000 instead of 10.

Procedure.—The experiment was conducted in Cambridge Colleges. One *E* canvassed for volunteers and kept curious eyes off the display. When *O* sat down at the table, the second *E* handed him a white card 8 × 5 in. with instructions typed on it, a pencil, and a white paper answer slip 4 × 3 in. Typical instructions for judgments of darkness read: "You will see two pieces of paper lying side by side, one medium gray and the other dark gray. If the darkness of the medium gray paper is called 10, what *larger* number describes the darkness of the dark gray paper? Please write down your answer on the slip of paper provided." The instructions were explained if necessary by paraphrasing them. For groups whose first variable had a greater reflectance than the standard, "light gray" was substituted in the instructions for "dark gray," and "smaller" for "larger." For groups judging lightness instead of darkness, "lightness" was substituted for "darkness," and "smaller" for "larger." One or 1,000 was substituted for 10 when required.

As soon as *O* had written down his first NE, the second *E* changed the variable and handed him an additional instruction card. The typical instructions which followed the example given above read: "What number *smaller* than 10 describes the darkness of the light gray paper? Please write down this second answer below your first." The procedure was repeated until a group of 50 acceptable pairs of NEs had been collected. Under 4% of the answers did not satisfy the instructions, and were therefore discarded.

Calculations.—Two distinct median measures were calculated. The *group median* is the median of the 50 *O*'s in a group. Owing to the tendency to select numbers ending in zero or five, occasionally the last 2 *O*'s could change the value of the group median by over 10%. The *mean median* is the average median of five subgroups, comprising the first and last 5 *O*'s in the group, the second and one from last 5 *O*'s, and so on (thus balancing for the position on the table of the standard relative to the variable). This mean median should be more stable, and has a standard error which permits *t* tests to be carried out. Two-tailed tests have always been used. The exponents (assuming that judged reflectance is a power function of physical reflectance), and their 5% fiducial limits were calculated from the logarithms of the five medians of each group.

RESULTS

The main results are shown in Tables 1 and 2. As far as the range of available reflectances would permit, the values of the variables were selected to produce predicted NEs of about 5 or 20, since very first NEs of about half the numerical value of the standard have been found to be less biased by *O*'s tendency to select a number a little too close to half (Poul-

TABLE 1

JUDGMENTS OF LIGHTNESS

Group	Standard	First Variable				Second Variable			
	% Reflectance	% Reflectance	Numerical Estimate			% Reflectance	Numerical Estimate		
			Group Median	Mean Median	Exponent		Group Median	Mean Median	Exponent
First Variable Lighter									
IIB	53	83	27.5	30.3	2.37±.99*	18.6	4	3.7	.96±.25*
IIIB	18.6	53	20	22.8	.76±.32	3.4	2	2.5	.84±.17
IIIIB	7.8	44	25.5	38.6	.78±.50*	3.4	4	3.9	1.15±.28
VB	.41	3.4	25	29.2	.49±.18*	.26	5	5.3	1.42±.60
VIB	.26	7.8	45	42.7	.41±.11	44	78	74.5	.39±.06
First Variable Darker									
ID	83	18.6	4	4.0	.62±.10	44	7	7.1	.55±.15
IID	53	18.6	5	4.7	.72±.08*	83	15	15.4	.96±.24*
IIID	18.6	3.4	1	1.6	1.15±.37	53	18	19.4	.63±.14
IIIID	7.8	3.4	4	4.1	1.08±.28	44	18	17.5	.32±.02*
VD	.41	.26	5	4.9	1.59±.54	3.4	18	18.5	.29±.03*

* With identical display, first and second judgments differ at $p \leqslant .05$.

TABLE 2

JUDGMENTS OF DARKNESS

Group	Standard	First Variable				Second Variable			
	% Reflectance	% Reflectance	Numerical Estimate			% Reflectance	Numerical Estimate		
			Group Median	Mean Median	Exponent		Group Median	Mean Median	Exponent
First Variable Lighter									
IIA	53	83	2	2.4	3.24±.58	18.6	30	33.9	1.13±.37
IIIA	18.6	53	4	3.6	.99±.15	3.4	18	17.9	.34±.05*
IIIIA	7.8	44	3	3.5	.62±.08*	3.4	16.5	16.3	.59±.08*
VA	.41	3.4	6	6.1	.24±.08*	.26	13	13.9	.70±.23
VIA	.26	7.8	5	4.5	.23±.04	44	2	2.1	.31±.02
First Variable Darker									
IC	83	18.6	101.7	125.7	1.67±.24	44	50	53.4	2.54±.76
IIC	53	18.6	25	31.5	1.06±.34	83	2.5	2.4	3.20±.58
IIIC	18.6	3.4	30.9	35.0	.72±.21*	53	3	3.1	1.12±.05
IIIIC	7.8	3.4	25.2	25.8	1.12±.29*	44	3	2.9	.72±.06*
VC	.41	.26	15.1	14.9	.87±.26	3.4	4	4.1	.42±.10*

* With identical display, first and second judgments differ at $p \leqslant .05$.

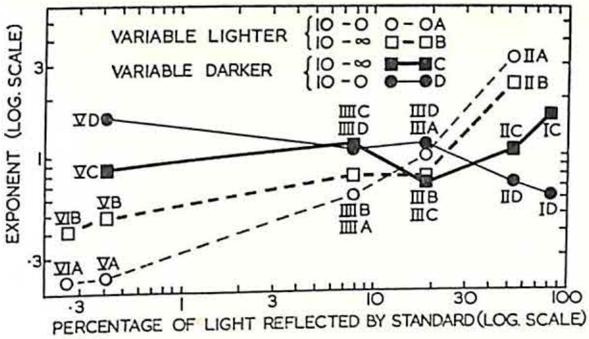


FIG. 1. Relationship between reflectance of standard and exponent calculated from the medians of the very first NEs of unpracticed *O*'s. (This is shown separately for a variable with a reflectance greater and smaller than the reflectance of the standard, and for a number range of from 10 to 0 and from 10 to infinity. Each point is from a separate group of 50 *O*'s, and is labeled to correspond to the labeling in Tables 1 and 2.)

ton & Webster, unpublished[2]; Stevens & Poulton, 1956). For Groups IIID and IIA who gave median first NEs of 1 and 2, respectively, it is possible that the exponent has been underestimated.

The exponents of the first judgments are plotted in Fig. 1. The reciprocals of the exponents are proportional to the differences in db. required for half or twice judged reflectance. Figure 1 is thus comparable in general orientation, though it is not related linearly in scale, to Stevens' (1955) Fig. 1, 2, and 6 which show the difference in db. required for half or twice loudness at different sound pressure levels. For the standards with reflectances of 83%, 52%, .41%, and .26%, all the differences between the points on the four functions are significant at the .05 level or better, but for the standards with reflectances of 18.6% and 7.8% only the more widely separated points are reliably different (see the 5% fiducial limits in Tables 1 and 2).

When the reflectance of the standard was high, variables with reflectances greater than the standard gave larger exponents than variables with reflectances smaller than the standard. The reverse was the case when the reflectance of the standard was low. The differences were particularly marked when the range of numbers available to *O* was restricted to between 10 and 0. The means of the points on the four functions for standards with reflectances of 53%, 18.6%, 7.8%, and .41% are, respectively, 1.85, .91, .90, and .80. If reciprocals are taken before the means are calculated, and the reciprocals of the means are then taken, the values are 1.31, .88, .86, and .50. This latter procedure corresponds to Stevens' (1955) method of computing the exponent from the average difference in db. of the physical magnitude required to give half or twice the sensory magnitude.

Contrast with background.—Poulton and Webster (unpublished)[2] found that when the reflectance of a gray background lay between the reflectances of the standard and variable, it exaggerated the size of the exponent. The points in Fig. 1 which may be too high as a result of the present brown background with a reflectance of 20% are IC and D, IIC and D, IIIA and B, and IIIIA and B. With the exception of IIIA and

B, these points are the lowest for their particular standards. IIIA and B have one point lying below them, but the difference is reliable statistically only in the case of IIIA ($p < .02$), whose position on the A function in Fig. 1 fits in neatly between the points IIA and IIIIA. Thus although an effect caused by the reflectance of the brown background cannot be ruled out, it is certainly not responsible for the general picture in Fig. 1.

Prior context.—For 16 groups of *O*'s the variable judged second lay on the opposite side of the standard to the variable judged first. These second NEs can be compared with the first NEs made with the same display by different groups. Tables 1 and 2 show that in eight instances (marked with asterisks) prior context affected the second judgment to a statistically reliable extent. The exponent of the second judgment of Group IIB in Table 1, which followed a first judgment using a range of numbers between 10 and infinity, is significantly larger than the exponent of the first judgment of Group IID. Similarly the exponents of the second judgments of Groups II, IIII, and VD, which followed a first judgment using a range of numbers between 10 and 0, are significantly smaller than the respective exponents of the first judgments of Groups II, IIII, and VB. There are two further instances of each kind in Table 2. They suggest that in attempting to make a second judgment consistent with the first, some *O*'s did not think in ratios; there was a tendency to look upon NEs of 5 and 15 as comparable with respect to the standard called 10.

Numerical value of standard.—Table 3 shows the four additional groups for whom the standard was called 1,000 or 1, each paired with the comparable group taken from Table 1 or 2 for whom the standard was called 10. Groups ID^3 and VIA^3 made fractional estimates with the standard called 1,000, and thus had available a

TABLE 3

NUMERICAL VALUE OF STANDARD

First Variable	Group	Standard		First Variable				Second Variable			
		% Reflectance	Numerical Value	% Reflectance	Numerical Estimate			% Reflectance	Numerical Estimate		
					Group Median	Mean Median	Exponent		Group Median	Mean Median	Exponent
				Judgments of Lightness							
Lighter	VIB^0	.26	1	7.8	6	5.5	$.49 \pm .06$	44	9.6	10.0	$.44 \pm .06$
	VIB	.26	10	7.8	45	42.7	$.41 \pm .11$	44	78	74.5	$.39 \pm .06$
Darker	ID^3	83	1000	18.6	350	363	$.69 \pm .17$	44	750	744	$.46 \pm .09$
	ID	83	10	18.6	4	4.0	$.62 \pm .10$	44	7	7.1	$.55 \pm .15$
				Judgments of Darkness							
Lighter	VIA^3	.26	1000	7.8	400	389	$.28 \pm .04$*	44	200	174	$.35 \pm .07$
	VIA	.26	10	7.8	5	4.5	$.23 \pm .04$*	44	2	2.1	$.31 \pm .02$
Darker	IC^0	83	1	18.6	10	11.1	$1.58 \pm .28$	44	6	5.7	$2.70 \pm .44$
	IC	83	10	18.6	101.7	125.7	$1.67 \pm .24$	44	50	53.4	$2.54 \pm .76$

* VIA^3 – VIA; $p < .05$.

greater range of numbers than Groups ID and VIA. Of the four differences between pairs of exponents, only the difference between the first judgments of Groups VIA[3] and VIA was significant ($p < .05$). It was in the expected direction of a higher exponent with the greater range of numbers, but the ratio was only 1:1.2.

Groups VIB° and IC° made multiple estimates with the standard called 1, and could thus select a number many times as large as the standard without the number appearing excessively large as it might have done for the comparable Groups VIB and IC. Table 3 shows that this did not produce a consistent or statistically significant difference. Thus with unpracticed *O*'s the exact choice of number for the standard is not as critical as it appears to be with practiced *O*'s who have associated particular numbers with particular sensory magnitudes. With experienced *O*'s judging loudness Hellman and Zwislocki (1961, Fig. 9) found that changing the number allocated to the standard by a factor of 100 changed the median estimated fractional ratio by a factor of up to about 10, and the corresponding multiple ratio by a factor of up to about 2.

DISCUSSION

Figure 1 shows that in the absence of the possible biasing effect of prior context (Garner, 1954, 1959; Poulton & Webster, unpublished[2]; Warren & Poulton, 1960), the exponent calculated from magnitude judgments of relative reflectance depends upon *E*'s choice of the reflectance of the standard and first variable—or presumably of the first two variables if no standard is present (Stevens, 1956, p. 19)—and on the range of numbers available to *O*. It appears to be impossible to obtain a direct measure of sensory magnitude which is unaffected by the experimental conditions under which it is made. This raises serious doubts as to whether it is possible to obtain direct measures of sensory magnitude having any general meaning, as opposed to a restricted operational meaning.

The finding that the physical magnitude of the standard is a determinant of the exponent when prior context is eliminated, appears to present considerable difficulties in comparing exponents derived from different sensory continua (Stevens, 1960a, 1960b, 1961a, 1961b). For it does not seem possible to state a priori what intensity of standard in one continuum should correspond to a given intensity of standard in another. Unless standards in different continua can somehow be equated, and these standards are used in the defining experiments with untrained *O*'s, it is clearly incorrect to ascribe the differences in exponent found with different sensory continua to the nature of the sensory "transducers" as Stevens (1960b, p. 251; 1961a, p. 27) has done. The range of exponents which Stevens has listed for different sensory continua, from .33 for brightness with a 5° target to 3.5 for a 60-cps electric shock through the fingers, corresponds remarkably well to the range of exponents of from $.23 \pm .04$ to $3.24 \pm .58$ found by us for the single continuum of reflectance by altering the conditions of measurement. Contrary to what has been claimed (Stevens, 1960b, p. 246; 1961a, p. 24) "cross-modality matching" does not validate the methods used to measure sensory magnitudes; it only demonstrates that experienced *O*'s (Stevens, Mack, & Stevens, 1960, p. 61) can learn to judge consistently in and between different sensory continua.

The optimist who still believes in the generality of direct measures of sensory magnitude might try averaging the four complete sets of data in Fig. 1 for different reflectances of the standard. A simple averaging of the four exponents gives means ranging from 1.85 for the standard of 53% reflectance, to .80 for the standard of .41% reflectance, a ratio of 2.3 to 1. Stevens' (1955) method of averaging the differences in db. required for half or twice the sensory magnitude gives means ranging from 1.31 to .50, a ratio of 2.6 to 1. These ranges of exponent for a single sensory continuum are too large to tolerate with intellectual comfort.

A more hopeful expedient would be to discard the data derived from standards which give the most widely different exponents for fractional and multiple judgments, on the grounds that these are effects related to one or other end of the physical range of stimuli, and are therefore "atypical." It would leave the data in Fig. 1 only for the standards of 18.6% and 7.8% reflectance. Using Stevens' technique for averaging, an exponent of .87 would be obtained. A practical disadvantage of adopting this expedient for all sensory continua is that it would first be necessary to run separate groups of inexperienced *O*'s with standards of representative physical magnitudes, and with first variables greater than and less than the magnitudes of the standards, in order to discover where the "typical" range of standards lay. This would require a large number of naive *O*'s, but is quite feasible.

REFERENCES

GARNER, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48**, 218–224.

GARNER, W. R. The development of context effects in half-loudness judgments. *J. exp. Psychol.*, 1959, **58**, 212–219.

HELLMAN, R. P., & ZWISLOCKI, J. Some factors affecting the estimation of loudness. *J. Acoust. Soc. Amer.*, 1961, **33**, 687–694.

KÜNNAPAS, T. M. Scales for subjective distance. Report No. 85, 1960, University of Stockholm, Psychological Laboratory.

STEVENS, J. C., MACK, J. D., & STEVENS, S. S. Growth of sensation on seven continua as measured by force of handgrip. *J. exp. Psychol.*, 1960, **59**, 60–67.

STEVENS, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, **27**, 815–829.

STEVENS, S. S. The direct estimation of sensory magnitudes-loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

STEVENS, S. S. On the new psychophysics. *Scand. J. Psychol.*, 1960, **1**, 27–35. (a)

STEVENS, S. S. The psychophysics of sensory function. *Amer. Scientist*, 1960, **48**, 226–253. (b)

STEVENS, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication*. New York: Wiley, 1961. Ch. 1. (a)

STEVENS, S. S. Toward a resolution of the Fechner-Thurstone legacy. *Psychometrika*, 1961, **26**, 35–47. (b)

STEVENS, S. S., & POULTON, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.*, 1956, **51**, 71–78.

STEVENS, S. S., & STEVENS, J. C. The dynamics of visual brightness. Report No. PPR-246, 1960, Harvard University, Memorial Hall.

STRANGERT, B. A validation study of the method of ratio estimation. Report No. 95, 1961, University of Stockholm, Psychological Laboratory.

TORGERSON, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications*. New York: Wiley, 1960. Ch. 3.

WARREN, R. M., & POULTON, E. C. Basis for lightness-judgments of grays. *Amer. J. Psychol.*, 1960, **73**, 380–387.

(Received April 27, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 305–308

EFFECT OF STIMULUS-RESPONSE MEANINGFULNESS ON PAIRED-ASSOCIATE LEARNING AND RETENTION

V. K. KOTHURKAR
University of Poona

Ss learned, recalled, and relearned 4 differently constituted paired-associate lists, consisting of dissyllables and paralogs of high and low meaningfulness. Both stimulus and response meaningfulness facilitated learning though the effect of the latter was much greater. As expected, the increasing rank order of difficulty of High-High, Low-High, High-Low, and Low-Low was obtained for learning; however, this hierarchy of effects was disturbed in the case of retention by Low-High scoring over even High-High. High meaningful responses are better remembered when paired with low meaningful stimuli than when associated with high meaningful stimuli. Inferior retention of High-High was explained in terms of spontaneous recovery of older connections and interference from them.

This study reports an experiment in which an attempt has been made to investigate the relative influence of the meaningfulness of the stimulus and response members on paired-associate learning and retention. Earlier studies have confined themselves to paired-associate learning alone. They generally explored the difficulty-meaning relationship for acquisition of verbal materials. It was assumed, that the differential effects of stimulus-response meaningfulness on paired-associate learning would be replicated for retention as well. In the present study, an attempt was made to explore not only the acquisition, but also the retention of paired associates, as determined by stimulus-response meaningfulness.

However, it was thought that the difficulty-meaning relationships in paired-associate lists could not be entirely independent of interference or transfer effects—"transfer gradients" as Woodworth and Schlosberg (1954, p. 754) have called them—from previous stimulus-response habits; there would be some confounding of effects. The transfer effects may be different according as new responses are made to the old stimulus units, or the same responses are made to new stimulus units, or both units are new. Difficulty or ease of paired-associate learning and retention is regarded as being a joint function of transfer gradients from previous stimulus-response relationships and the inherent meaningfulness of stimulus and response terms. Actual ease or difficulty of learning and retention of variously constituted paired-associate series would depend upon a delicate balance of transfer gradients and the facilitative effects of meaningfulness.

Again, it is also conceivable that the interaction of meaning and transfer effects may be different for learning and retention of paired associates.

In the context of these considerations, the following hypotheses were formulated:

1. In learning High-High series, there would be some interference from past associations of the high meaningful stimuli. However, this interference would be counteracted by the present response set and by the facilitative effect of response meaning on the learning of paired associates.

2. In the case of retention (as measured by recall, saving, etc.) there would be greater competition and interference from the spontaneous recovery of old associations of high meaningful stimuli, as these would not be inhibited by any present response set. This greater interference would not be adequately counteracted or compensated by response meaningfulness and, therefore, meaningful response members would be better remembered when they are paired with low meaningful stimuli. In short, it was hypothesized that the High-High would be superior to Low-High in learning. However, this advantage may be reversed in retention and Low-High may actually be better remembered than High-High.

METHOD

Design of the experiment.—A simple 2 × 2 factorial design was employed in the learning part of this experiment; and it followed the typical procedure of the paired-associate method in the study of learning as well as retention.

Materials.—Four paired-associate lists, each consisting of 10 pairs taken from Noble's (1952) list of dissyllables calibrated for meaning (*m*), constituted the materials of this experiment. High-High, Low-High, High-Low, and Low-Low lists were constructed by appropriate combinations of the scaled dissyllables and paralogs. The average *m* values of the stimulus and response sides of the lists were High (7.62)-High (7.03), Low (1.25)-High (7.03), High (7.03)-Low (1.25), and Low (1.25)-Low (1.29).

A practice list of four pairs was used to acquaint the *S*s with the apparatus and the paired-associate procedure.

Subjects.—In all, 320 undergraduate and postgraduate students participated in this experiment on a voluntary basis. The mean age of the group was 21 yr.

Procedure.—A memory drum presented first the stimulus word of each pair, for 3 sec., during which the *S* was required to anticipate and produce the associated response member, and then the stimulus and response words of the pair together were presented for another 3 sec. The *S*s learned by the anticipation method except for the first two learning trials. They pronounced the dissyllables in an agreed, uniform way. The pairs were presented in three randomized sequences, as a precaution against possible position bias. There was a rest interval of 15 sec. after each trial, including the first two runs through the list. Standard paired-associate instructions were given to the *S*s.

The 320 *S*s were randomly assigned, in groups of 80, to learn the lists HH, LH, HL, and LL. In order to study retention, each group of 80 *S*s was further divided into four subgroups of 20. The four subgroups took their retention tests 24 hr., 48 hr., 96 hr., and 1 wk. after the learning session. Learning was measured in terms of the number of trials required to reach the criterion of one errorless recital. Retention was measured by computing: (*a*) the number of items recalled after specified interval, and (*b*) the percent saving in relearning the list to the same criterion.

Learning.—The learning scores of the experimental groups [illegible] 12.55, 14.1[illegible] [illegible]3, and 33.7[illegible] [illegible] High-High, Low-High, High-Low, and Low-Low lists, respect[illegible]. The results clearly show an increasing rank order of difficulty for the HH, LH, HL, and LL lists. These results are generally in agreement with those obtained by Cieutat, Stockwell, and Noble (1958), and Hunt (1959) and the two predictions of Noble and McNeely (1957, p. 21). The list with high meaningful stimuli and high meaningful responses is definitely easier than any other list. Correspondingly, the list with low meaningful stimuli and low meaningful responses is the most difficult. Similarly, as expected, the effect of response meaning is greater than that of stimulus meaning. Also, the difference between the mean scores for the two High-R lists and the two Low-R lists is greater than the difference between the mean scores of the two High-S and the two Low-S lists.

However, whereas Hunt (1959) reports that "Variations in response meaningfulness produced a significant effect upon acquisition. . . . Stimulus meaningfulness and articulation produced no significant effects" (p. 266), in the present experiment, as the analysis of variance in Table 1 reveals, *F*s for response meaningfulness as well as stimulus meaningfulness are highly significant, having probabilities of less than .01.

Thus, under the conditions of the present experiment, either of the following inferences are indicated: (*a*) lists with high meaningful response terms are significantly easier than the lists with low meaningful response terms, and (*b*) lists with high meaningful stimulus terms are also significantly easier than the lists with low meaningful stimulus terms.

There is a significant interaction between the two variables, response meaning and stimulus meaning, so that (*a*) the difference in learning scores between the high response-meaning lists and the low response-meaning lists is less when stimulus members are highly meaningful than when they are of low meaningfulness, and, similarly, (*b*) the difference in learning scores between the high stimulus-meaning lists and the low stimulus-meaning lists is less when response members are highly meaningful than when they are of low meaningfulness.

TABLE 1

ANALYSIS OF VARIANCE OF LEARNING SCORES

Source	*df*	*F*
Response meaning (R)	1	158.11**
Stimulus meaning (S)	1	119.96**
R × S	1	87.37**
Within groups (error *MS*)	316	(75.46)

** $p < .01$.

TABLE 2

MEAN RECALL SCORES OF THE EXPERIMENTAL GROUPS AT DIFFERENT TIME INTERVALS

Time Intervals	Experimental Conditions			
	High-High	Low-High	High-Low	Low-Low
24 hr.	8.55	9.00	8.55	6.90
48 hr.	8.45	8.80	7.10	7.50
96 hr.	7.65	9.00	7.35	6.75
1 wk.	7.15	9.10	6.50	6.80
Means for the principal experimental groups	7.95	8.98	7.38	6.99

Note.—Each cell mean is based on 20 recall measures.

TABLE 3

ANALYSIS OF VARIANCE OF RECALL SCORES

Source	*df*	*F*
Time intervals (T)	3	3.83[a]
Experimental conditions (E)	3	59.78**
T × E	9	1.81
Within sets (error *MS*)	304	(2.86)

Note.—Time intervals involved are 24 hr., 48 hr., 96 hr., and 1 wk.

[a] *F* is almost significant at .05 level.

** $p < .01$.

Bartlett's test showed unequal variances among the four experimental groups. However, these do not invalidate the very highly significant main effects. Moreover, Bartlett's test brought out an inverse relationship between ease of learning a list and variability in mastering it.

Retention.—The mean recall scores (number of items recalled) of High-High, Low-High, High-Low, and Low-Low were 7.95, 8.98, 7.48, and 6.99, respectively. The High-High, Low-High, High-Low, and Low-Low hierarchy of effects, obtained for learning was, however, disturbed, in the case of retention, by Low-High scoring over High-High.

Analyses of variance in two criteria classifications reveal that variations in time interval, and more importantly, variations in experimental conditions of stimulus and response meaningfulness carry with them systematic variations in recall scores. Mean recall scores and their analysis of variance are presented in Tables 2 and 3, respectively.

Maximum saving occurred in LH and minimum in HH, for all time intervals, without any reversions. However, no tests of significant differences were made between pairs.

As was hypothesized, it is important to note that, in retention as measured by recall, the experimental condition Low-High is superior not only to its reverse High-Low, but also to High-High and Low-Low.

This superiority of the experimental condition Low-High over even High-High needs to be further considered.

DISCUSSION

The main results of this experiment are consistent with previous work. Meaning generally facilitates learning and retention. It is also more helpful in the response position. The expected hierarchy of effects was observed in learning. However, there was deviation from this hierarchy in the direction of our hypothesis in retention.

In learning HH some interference or negative transfer from old multiple associations of the meaningful stimuli is to be expected. However, as learning proceeds, the experimental responses alone are reinforced and the extraneous associations, not being reinforced, are extinguished and inhibited.

This extinction or inhibition of old S-R habits, generated in the learning of HH, is dissipated during the interval between learning and recall or relearning tests. The extinction process, which blocks and neutralizes habit strength of old associations, weakens by passage of time, thus allowing old connections to interfere with the recall of experimental response units. There is some spontaneous recovery of old connections which now interfere and reduce the recall score of HH as compared with LH.

According to Hull's theory, performance is dependent on habit strength and present drive. In learning HH, old habits are held in abeyance and the present drive favors the production of reinforced responses.

However, in remembering, the individual is thrown back upon his past, learned reactions and associations; there is competition and rivalry among them, and thus causing greater interference and reduced recall. In LH, the low meaningful stimulus has fewer or no associations connected with it, and hence there is less interference and better recall. By the time of the retention test, the old inhibited bonds spontaneously recover and, because of the relative freedom from the bounds of the present, cause greater interference in the recall of HH.

The superior retention of LH, as compared with HH, may also be accounted for in terms of freedom from interference and isolation effects. Under HH, mutual interference of other associations, both from within the stimulus and response sides of the list and from outside the list, result in inferior recall; in LH, the stimulus side consisting of meaningless paralogs provides an excellent interference-free background against which the meaningful response members stand out sharply and unambiguously, and, hence, there is better retention.

REFERENCES

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (*m*, *m'*) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

HUNT, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, **57**, 262–267.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

NOBLE, C. E., & MCNEELY, D. The role of meaningfulness (*m*) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

(Received May 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 309–316

THE ORIENTING REFLEX AS A FUNCTION OF THE INTERSTIMULUS INTERVAL OF COMPOUND STIMULI[1]

CHARLES K. ALLEN, FRANCES A. HILL, AND DELOS D. WICKENS

Ohio State University

The experiment investigated the orienting response as a function of temporal relations among the stimuli, with GSR the measure. 3 stimuli were employed. The duration of the 1st was 1300, 1050, or 750 msec., with 2 groups at each interval. The duration of the 2nd and 3rd were 600 and 100 msec., respectively. They terminated together and consisted of a tone and 2 different lights. Following 10 presentations of the compound, separate groups were tested to the 1st or 2nd stimulus alone. The shift from triple stimulation to single produced a significant increase in GSR magnitude and also a significant interaction between order and original precedence time.

In 1959 Wickens, Gehman, and Sullivan published an article which dealt, in a specific situation, with the response strength accruing to the elements of a compound stimulus which has been associated with reinforcement. The situation with which they worked was classical GSR conditioning to shock reinforcement. The independent variable of the experiment was the time of onset of one of the CS elements of the compound with respect to the second CS element as well as to the UCS. Specifically, the two conditioned stimuli employed were a light and a tone and they were compounded in the following fashion: One of these stimuli always preceded the shock (UCS) by 500 msec., a temporal separation which has been presumed to result in optimal conditioning when a single stimulus serves as the CS (Moeller, 1954; White & Schlosberg, 1952). The other stimulus, the longer one of the pair, preceded the shock in the separate groups of the experiment by times ranging from 600 to 4600 msec. Following the 10 conditioning trials, the *S*s were tested to either the long or short element of the pair. Control groups, conditioned to a single stimulus, paralleled the interstimulus intervals selected for the longer element groups. The curve for these single stimulus groups showed the expected monotonic decline in response strength as the interstimulus interval increased from 500 msec. to approximately 1500 msec. The curve for the long element of the complex, however, rose to a maximum in the neighborhood of 900 msec. and then dropped rather sharply; and the curve for the shorter element—plotted against the time of onset of the longer element—can economically, and with adequate accuracy, be described as the mirror image of the long stimulus curve.

Since the date of publication of this article a rather considerable amount of research has been published on what has been variously called the orienting reflex (Razran, 1961; Sokolov, 1960) or the perceptual disparity reaction (Grings, 1960). Essentially,

---

[1] This research was supported by a National Science Foundation Grant, G 6203, to D. D. Wickens.

This experiment was submitted by the senior author as his MA thesis at the Ohio State University.

these concepts refer to the fact that if a particular stimulus is given for a number of trials during which series there is usually a continuous decrement in GSR magnitude on successive stimulations, and the series is immediately followed by a change in the nature of the stimulation, there will be a marked increment in the GSR magnitude. In short, novelty of stimulation is a sufficient condition for the evocation of the GSR. It is apparent that the element testing procedure employed in the Wickens, Gehman, and Sullivan experiment introduces the necessary conditions for the evocation of the orienting reflex and its associated GSR. It does so because the 10 pairings of light, tone, and shock are immediately followed by tests to a single element and this constitutes a disruption in the previous pattern of stimulation. If one assumes that the degree of perceptual disparateness generated by an element when presented alone is a function of its temporal relationship to the other stimuli in the complex, then it is possible to suppose that the curves generated by the elements in the Wickens, Gehman, and Sullivan study are a measure of the orienting response and not the result of some association established between the two conditioned stimuli and the shock.

The present experiment was undertaken to investigate this possibility. Essentially, it simply substituted for the shock UCS the innocuous illumination of a neon bulb. That is, it too employed three stimuli, presented in certain temporal relations, but the third did not have the consistent and powerful response evoking characteristics of the shock. Three different intervals between onset of the long and short element—150, 450, and 700 msec.—were employed. These intervals were chosen because relatively large differences in response strength to the long and short element would, according to the Wickens, Gehman, and Sullivan results, be expected at these particular times, with the short greater than the long at 150, the reverse being true at 450, and with another reversal at 700.

METHOD

Subjects.—The *S*s were 141 students from the elementary psychology course at the Ohio State University. The data for 120 *S*s are reported, the remainder being eliminated because of base resistance above 100,000 ohms, *E* errors, or equipment failure. There were 20 *S*s in each of the six groups of the experiment.

Apparatus.—Timing was controlled by Tektronix equipment consisting of three wave-form generators, two pulse generators, and a power supply. Stimulus duration and interstimulus intervals were measured by means of a Hunter Klockkounter, and according to these measurements, timing was accurate to ±3 msec. The auditory "CS" was a 1000-cycle tone at 60 db. (re 0002 dyne/cm^2) generated by a Jackson audio frequency oscillator and delivered through Willson sound barrier earmuff earphones. The visual CS was produced by the illumination of a white 2-w. lamp placed behind a 2-in. square white panel. It was located at eye level approximately 6 ft. directly in front of *S*. Two neutral Wratten filters reduced the intensity to a level which was judged by several *O*'s to be equal in psychological effectiveness to the tone. The UCS surrogate was produced by a General Electric Ne 40 neon bulb located 3½ in. below the visual CS. The GSR was measured by means of a Fels dermohmeter and recorded from electrodes clamped to the palm and back of the left hand. The ¾-in. diameter zinc electrodes were contained in plastic cups and they were freshly filled for each *S* with an electrode jelly consisting of zinc sulphate, bacto-agar, and distilled water. The *E* recorded the response visually to the nearest 200 ohms from a meter on the face of the instrument. A reading was taken of the resistance at the time the stimulus was administered (R_b) and of its minimal value within 2 sec. after stimulation (R_a). A score was obtained by subtracting the reciprocals of these resistances, multiplying by 10^8, adding 1.0, and obtaining the common log of this number.

The *S*'s room was adjacent to the experimental room and contained a ventilating fan that, together with the earphones and a partition between the rooms, prevented *S* from hearing the operating noise produced in *E*'s room. During the experiment *S*'s room was dimly illuminated by a 6-w. night light.

Conditions and groups.—Six different groups of *S*s were employed in the experiment and they differed from each other with respect to the temporal duration of the longer stimulus or to the element to which "extinction" tests were to be given following training to the compound. For all groups the shorter element began 500 msec. before the neon bulb flash (the shock surrogate) and terminated simultaneously with it. The duration of the bulb flash was 100 msec. The duration of the long stimulus, which was coterminous with the short stimulus and neon bulb flash, varied among the groups. For one pair of groups it preceded neon bulb onset by 650 msec., for another pair by 950 msec., and by 1200 msec. for the final pair. Thus, their times prior to onset of the short stimulus were 150, 450, and 700 msec., respectively.

Procedure.—The *S*s were assigned randomly to the treatment conditions, with the provision that the groups be nearly equal in size at the end of each week.

After *S* was seated, he was told that his responses were being measured to various stimuli and that he would hear a tone and see two lights, but that there was nothing which he need do other than to keep his eyes open and the hand on which the GSR electrodes were attached relatively still.

The "training" procedure in this experiment was an exact duplicate of the Wickens, Gehman, and Sullivan procedure, save that the light flash was used as a shock surrogate on these 10 trials. The procedure was also duplicated during the 5 "extinction" trials with a group at each interval being tested to the short stimulus and another group to the long stimulus. As in that experiment, the tone was the stimulus employed during this "extinction" series. These tone trials were followed by three similarly patterned tests to the CS light alone.

RESULTS

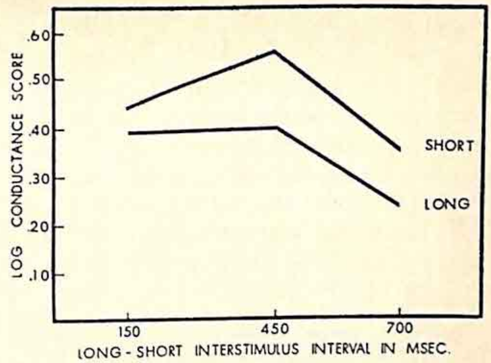


FIG. 1. Mean magnitude of response during the last two training trials for all groups.

Figure 1 shows the mean log conductance scores for the long and short stimulus groups for the last two acquisition trials as a function of time interval. The responses at this point in training were very small. In fact, of the total 120 *S*s, 75 gave no response on either Trial 9 or 10, 41 *S*s did not respond on both of the last two acquisition trials. Since the distribution of scores was so markedly skewed, the nonparametric Kruskal-Wallis one-way analysis of variance was computed for the long and short groups. The test indicated no significant differences for either the long or the short stimulus groups, and Mann-Whitney *U* tests between tone and light groups at each time interval were not significant.

To test whether *S*s responded significantly greater during the tests to

TABLE 1

ANALYSIS OF VARIANCE FOR DIFFERENCE SCORES: TESTS TO THE TONE

Source	*df*	*SS*	*MS*	*F*
Between-*S*	119	186.98765		
Long or short stimulus (B)	1	.01374	.01374	.009
Time interval (C)	2	.07197	.03598	.024
B × C	2	14.99610	7.49805	4.972**
Error	114	171.90584	1.50794	
Within-*S*	480	202.17614		
Disparity trial (A)	4	8.75307	2.18827	5.438**
A × B	4	1.47469	.36867	.916
A × C	8	7.15247	.89406	2.222*
A × B × C	8	1.31509	.16439	.409
Error	456	183.48082	.40237	
Total	599	389.16379		

* $p < .05$.
** $p < .01$.

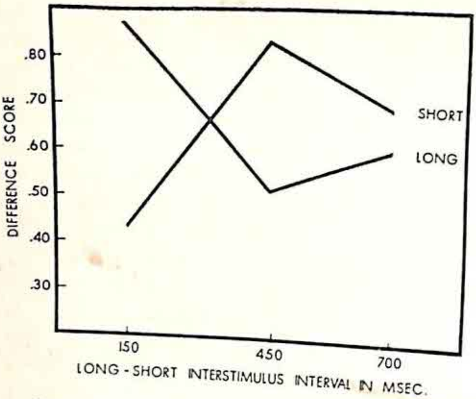


FIG. 2. Magnitude of the perceptual disparity response as a function of interstimulus interval during training and the element to which the test was made.

the single stimulus, hereafter referred to as the disparity trials, than during the last two "acquisition trials," the mean of these two acquisition trials was computed for each *S* and this mean was subtracted from each *S*'s score on the disparity trials. Wilcoxon matched-pairs signed-ranks tests were computed for all groups on all disparity trials. For the first five disparity trials—the tests to the tone alone—all groups responded significantly higher on all the disparity trials than on the last two training trials with two exceptions. The exceptions were Trial 2 for the 450 long-tone group and Trial 5 for the 150 short-tone group.

TABLE 2

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES: TESTS TO THE LIGHT

Source	*df*	*SS*	*MS*	*F*
Between-*S*	119	128.13830		
Long or short stimulus (B)	1	.26744	.26744	.243
Time interval (C)	2	.04973	.02486	.022
B × C	2	2.65236	1.32618	1.208
Error	114	125.16878	1.09797	
Within-*S*	240	121.30353		
Disparity trial (A)	2	11.11372	5.55686	12.185**
A × B	2	.09011	.04506	.098
A × C	4	3.38417	.84604	1.855
A × B × C	4	2.73755	.68444	1.501
Error	228	103.07778	.45604	
Total	359	249.44183		

** $p < .01$.

Figure 2 shows the mean of the difference scores for all of the disparity trials to the tone alone. An analysis of variance was done on these data and the summary table is shown in Table 1. The main effects of time interval and long or short stimulus were not significant, but the effect of trials reached the .01 level. In addition, the interaction of time interval and long or short stimulus was significant at the .05 level. The source of the interaction of time interval and the long or short stimulus is rather obvious from an inspection of Fig. 2. A *t* test indicated that the long stimulus group is significantly superior ($.05 > p > .01$) to the short stimulus group at the 150 interval, while the short is significantly superior to the long at the same level of confidence at the 450-time interval. At the 700 interval the long and short stimulus groups did not differ significantly.

The same analysis was done on the disparity trials to the light alone, and a summary of this analysis is shown in Table 2. The only significant effect at this point in the experiment was that of trials. The *t* tests between

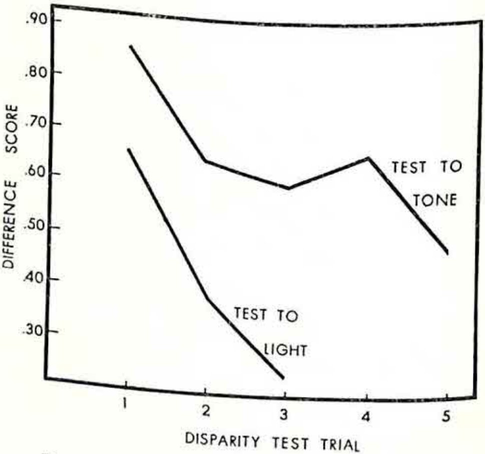


FIG. 3. Extinction of the perceptual disparity response.

the long and short stimulus groups were not significant at any of the three intervals. A Wilcoxen matched-pairs signed-ranks test between the mean of Trials 9 and 10 and the first test to the light was significant in every group, but no significant differences were found for any groups on the third trial.

Figure 3 shows the disparity effect for the tests to the tone and tests to the light, respectively, as a function of trials. The data are the mean difference scores for all of the groups at each disparity trial and they clearly show a progressively smaller response with continual presentations of one element of the complex alone. In each analysis this accounts for the significant effect of trials. A *t* test comparing the last tone trial with the first light trial was significant between the .02 and .01 levels.

DISCUSSION

This experiment was undertaken to determine whether or not the results obtained in the Wickens, Gehman, and Sullivan experiment could be accounted for by an orienting reflex or perceptual disparity reaction consequent to the procedural change involved in testing to a single element. Essentially this experiment asked two questions, the first was whether or not an increased GSR would occur following the procedural change, and the second was whether or not its magnitude would show the same functional relationship to the temporal variable of the CS_1-CS_2 relationship found by Wickens, Gehman, and Sullivan. The CS_1-CS_2 difference values of 150, 450, and 700 msec. were chosen because the value for the long stimulus at these intervals would be expected to be low, high, and low, respectively. The opposite relationship should hold for the short stimulus, namely high, low, and high.

The results of the experiment confirm the expectation that a perceptual disparity reaction or orienting reflex does occur when the element testing procedure is introduced. The magnitude of the response is increased significantly over that which was obtained on the final two trials of the "training" series, and this response increment decreased upon continued presentation of the stimulus. These are two essential characteristics of the orienting reflex as described by Sokolov (1960). The habituation of the orienting reflex is described in Fig. 3.

The rather rapid decrease in the amount of disparity reaction when tests were made to the light alone is consistent with results obtained by Coombs (1938). This investigator presented tones to an *S* while recording GSR. He found that the response decreased with continual presentations of the stimulus, but that the response returned when the stimulus was changed. When the changed stimulus was repeatedly presented, Coombs found that the response was habituated more rapidly than was the case with the original stimulus. The test trials in the present experiment seem to be analogous to the Coombs experiment. After habituation to the tone, the stimulus was changed to the light. An increase in response was noted when the light was presented, but the orienting reflex was habituated after only three trials.

Thus, the results of the experiment clearly indicate that a GSR is generated by the shift in procedure involved in testing to an element of the compound stimulus, and that it persists with declining strength for at least five trials. In addition, the magnitude of this response proved to be an interacting function of order of presentation of the elements of the compound and their interstimulus interval in that the long and short elements differ significantly from each other at the interstimulus intervals of 150 and 450 msec. The long produced the larger response at 150 and the short at 450 msec. Presumably these reflect the operation of a perceptual disparity reaction or the orienting reflex. If shock is used as the UCS, as in the Wickens, Gehman, and Sullivan experiment, significant differences in response strength to the two elements at these interstimulus intervals are also expected. Under those conditions, however, the magnitude relationships are reversed, with the long stimulus evoking the smaller response at 150 and the larger response at 450 msec.

Although the curves for the long and short stimulus elements across time in this experiment are essentially the opposite of those obtained in the Wickens, Gehman, and Sullivan experiment, the interpretation which was used to explain the earlier curves seems to be applicable in the present case.

This theory might be called a sensory conditioning-stimulus generalization interpretation. It stated that the actual CS in the shock experiment is the interacting sensory effect of the two elements, and the response which is evoked by an element is a function of its ability to reinstate in the organism—through a process of sensory conditioning—the sensory effects of the compound. Using the concept of stimulus generalization, it stated that the more similar the effect of the single stimulus to the effect of the compound, the greater the response evoked by that element. The interpretation further assumed that the effectiveness of the sensory conditioning will be a function of the time interval between onset of CS_1 and CS_2 (Hoffeld, Thompson, & Brogden, 1958; Silver & Meyer, 1954), and that this function has the same form as does the interstimulus curve obtained from work with a single stimulus (Wickens & Cross, 1960).

The application of this interpretation to the present experiment is as follows: One or another element, as a function of the time interval, becomes to a greater or lesser degree the dominant or salient element of the complex. This would be true of the short element at the 150-msec. interval because it is located at a near optimal interval—500 msec.—with respect to the light flash, whereas the long stimulus precedes it by the less effective time interval of 650 msec., but at a CS_1-CS_2 interstimulus interval too short for effective sensory conditioning. At 450 msec. the long stimulus dominates because it is located at a near optimal interval before the short stimulus so far as sensory conditioning is concerned, and, therefore, evokes the characteristics of the complex itself. The short stimulus, however, because of its backward relationship to the long stimulus is unable to elicit the characteristics of the compound. At the 700 interval the long stimulus is too distant in time to be a highly effective sensory conditioned stimulus for either the short stimulus or the light flash and thus the short stimulus—still 500 msec. from the light flash—begins to dominate the stimulus complex. In summary, the dominant stimulus of the complex is the short, the long, and possibly once again the short at the 150, 450, and 700 intervals, respectively.

If the dominant stimulus of the complex is presented alone, its occurrence—simply because it has dominated the complex—should not produce a marked disparity reaction. That is, this stimulus, by itself, has acquired the capacity to produce an effect quite similar to that of the compound. There is, then, for *S* no great disparity between this event and those which have occurred before, and hence, only a small perceptual disparity reaction or orienting reflex results. The converse would be true for the minor stimulus when presented alone. Its single occurrence is perceived as being markedly different from the prior stimulation and hence, a strong perceptual disparity response occurs. Essentially this is what the data show to be the case. The dominant short stimulus at the 150 interval produces a significantly smaller response than does the minor long stimulus; and the long stimulus, now dominant at the 450 interval, is significantly less effective in response evocation than is the short, or minor, component of the complex. At the 700 interval there is no significant difference between the two points and thus the prediction is not perfect. With respect to this temporal point, it should be added that the information concerning the courses of the long and short stimulus curves at these longer intervals is not clearly established.

This interpretation assumes that the

magnitude of the GSR would be a positive function of the magnitude of the perceptual disparity reaction, and this interpretation seems to be consonant with perceptual disparity theory. In describing the perceptual disparity response, Grings (1960) remarks:

> It is defined as the difference in magnitude of response between situations where receipt of stimulation following a signal cue is in accord with past experience and where receipt of stimulation is not in accord with past experience with the particular cue (p. 251).

By quoting an experiment on stimulus generalization performed by Kimmel (1960) he makes it apparent that he also assumes that the magnitude of the evoked response is positively related to the magnitude of the disparity.

It would appear, then, that the results obtained by Wickens, Gehman, and Sullivan cannot be interpreted as being solely the consequence of the orienting reflex, even though it does appear that the magnitude of the orienting reflex is a function of the temporal relationship of the elements of the complex. The Wickens, Gehman, and Sullivan results seem to be dependent upon the occurrence of a response-eliciting UCS during the presentation of the compound. Actually, the manner in which the orienting reflex in the element testing situation operates is probably to minimize differences, for its effect produces responses which are opposite in direction to the shock associated curves. It is suggested, however, that the same mechanism—sensory conditioning—operates in the same way in both situations.

When the light alone was presented following the five "extinction" trials to the tone, orienting reflexes or perceptual disparity reactions were present, as evidenced by a significant difference in magnitude of response compared with the last two training trials. There was, however, no significant interaction term as in the tests to the tone. Although at first blush it might appear that such an interaction is again to be expected, it should be pointed out that the perceptual disparity responses in this session would be a function not only of the initial training, but also of the five tests to the tone alone. As a consequence of this confounding, it is difficult to predict what the effect of the original differences in time interval may have upon response magnitude in this situation.

Regardless of the validity of the above interpretation, this experiment, along with the article by Grings (1960), adds a note of warning for research workers in the field of GSR conditioning. It gives evidence that when procedural changes are introduced—as occurs when one goes from training to extinction—responses are evoked by the changes in the procedure itself, and the effect of this procedural shift persists for at least several trials. It is very likely that this effect could mask that of some other variable for at least the first few trials under the new condition. Quite possibly the flat gradient of stimulus generalization which is typically found on the first test trial in GSR conditioning (Hovland, 1937; Wickens, Schroder, & Snide, 1954) could be so interpreted.

REFERENCES

COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *J. exp. Psychol.*, 1938, **22**, 244–268.

GRINGS, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychol. Rev.*, 1960, **67**, 243–252.

HOFFELD, D. R., THOMPSON, R. F., & BROGDEN, W. J. Effect of stimuli time relations during preconditioning training upon the magnitude of sensory preconditioning. *J. exp. Psychol.*, 1958, **56**, 437–442.

HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 279–291.

KIMMEL, H. D. The relationship between direction and amount of stimulus change and amount of perceptual disparity response. *J. exp. Psychol.*, 1960, **59**, 68–72.

MOELLER, G. The CS-UCS interval in GSR conditioning. *J. exp. Psychol.*, 1954, **48**, 162–166.

RAZRAN, G. H. S. The observable unconscious and the inferable conscious in current

Soviet psychophysiology: Interoceptive conditioning, semantic conditioning, and the orienting reflex. *Psychol. Rev.*, 1961, **68**, 81–147.

Silver, C. A., & Meyer, D. R. Temporal factors in sensory pre-conditioning. *J. comp. physiol. Psychol.*, 1954, **47**, 57–59.

Sokolov, E. N. Neuronal models and the orienting reflex. In Mary A. B. Brazier (Ed.), *The central nervous system and behavior*. Madison, N. J.: Madison Printing, 1960.

Wickens, D. D., & Cross, H. A. Resistance to extinction as a function of temporal relations during sensory preconditioning. *Amer. Psychologist*, 1960, **15**, 494.

Wickens, D. D., Gehman, R. S., & Sullivan, S. N. The effect of differential onset time on the conditioned response strength to elements of a stimulus complex. *J. exp. Psychol.*, 1959, **58**, 85–93.

Wickens, D. D., Schroder, H. M., & Snide, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–55.

White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

(Received May 4, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 317

SUPPLEMENTARY REPORTS

A REPLICATION OF OVERLEARNING AND REVERSAL IN A T MAZE[1]

WINFRED F. HILL AND NORMAN E. SPEAR
Northwestern University

Rats were trained to criterion in a T maze and then given reversal training either with or without 140 trials of overtraining (16 Ss per group). The overtrained group required significantly more trials to reach the reversal criterion.

Hill, Spear, and Clayton (1962) recently reported three failures to find facilitation of T maze reversal by overlearning. Because they used a fixed number of trials instead of a criterion both for acquisition and for reversal, and because the control *S*s received extra handling to match that of the overtrained *S*s, these experiments were not entirely comparable to earlier ones in which overtraining has facilitated reversal. The present experiment was designed to see whether, when these differences from other studies were eliminated, an overlearning-reversal effect would appear. In particular, an attempt was made to match Pubols' (1956) procedure in his Exp. 2 as closely as possible, since his study also used a spatial discrimination.

Method.—The *S*s were 32 experimentally naive female albino rats, 74–81 days old at the beginning of training. They were divided randomly into two groups of 16, corresponding to Groups N and Fr in Exp. 2 of the previous study (Hill et al., 1962). The apparatus and procedure were the same as for those groups in the previous experiment, with the following exceptions. (*a*) All *S*s received 20 trials a day throughout all stages of training. For recording purposes, each day was regarded as consisting of two blocks of 10 trials. If *S* reached criterion on the first block of a day, the second block was omitted. (*b*) Original acquisition was carried to a criterion of 18 correct in 2 successive blocks of 10 trials, with 9 of the last 10 correct. This meant that the number of trials to criterion had to be a multiple of 10, the purpose being to make this study comparable to Pubols' (1956) Exp. 2. (*c*) Group N received only the same 6 days of prehandling as Group Fr. (*d*) Group Fr received 140 trials of overtraining. (*e*) A food cup (with or without food) was present in both end boxes on all trials. (*f*) Reversal training was carried to the same criterion as in acquisition.

Results.—The mean number of trials to the beginning of the acquisition criterion was 15.00 for Group N and 18.75 for Group Fr. These values were not significantly different (*t* of 1.10).

In reversal, Group N required a mean of 13.75 precriterial trials, while Group Fr required 23.12. This difference was significant at the .01 level (*t* of 3.16).

The mean total speed (based on time from the starting treadle to one of the two arm treadles) on the first five trials of reversal was greater for Group Fr than for Group N. The difference fell well short of significance, however, with a *t* of 1.21.

Discussion.—These results confirm the previous finding, significant in two of the three experiments, that T maze reversal is slower following overlearning. Thus the previous results of Hill, Spear, and Clayton cannot be explained by their use of a fixed number of trials in acquisition and reversal or by the extra handling of their control animals.

The trend toward faster running by the overtrained *S*s at the beginning of reversal is the same as that found in all three experiments of the previous study. However, in only one of these four cases was the difference statistically significant.

REFERENCES

HILL, W. F., SPEAR, N. E., & CLAYTON, K. N. A comparison of T maze reversal learning after several different overtraining procedures. *J. exp. Psychol.*, 1962, **64**, 533–540.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

(Received April 7, 1962)

---

[1] This research was supported by Grant G-8706 from the National Science Foundation.

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 318–319

EFFECT OF TRANSITIONS FROM NONREINFORCED TO REINFORCED TRIALS UNDER SPACED-TRIAL CONDITIONS

E. J. CAPALDI AND PATRICIA WARGO

University of Texas

2 groups of 11 rats each were trained to traverse a straight alley under partial reinforcement, the intertrial interval being 20 min. in acquisition and extinction. For Group SA (single alternation) the number of transitions from nonreinforced to reinforced trials was greater than for Group R (random training). Group SA was considerably more resistant to extinction than Group R. Since these results are consistent with deductions available from the Hull-Sheffield hypothesis, it was concluded that they offer a basis for assuming that aftereffects remain functional for at least 20 min.

A previous experiment (Capaldi & Hart, 1962) demonstrated that when two patterns of partial reinforcement involved the same small number of trials and the same number of reinforcements and nonreinforcements, the pattern which involved the greater number of transitions from nonreinforced to reinforced trials was followed by the greater resistance to extinction. The intertrial interval was 15 sec. The purpose of this experiment was to determine if similar results could be obtained under conditions of considerable trial spacing.

Method.—The *S*s were 22 experimentally naive albino rats. The straight-alley runway was identical in all respects to that employed by Capaldi and Hart (1962) except that it was 6 in. longer and its sides were 1 in. higher. Preliminary and experimental training procedures were identical to those employed in the earlier experiment except for the following changes: there were 5 additional days of preliminary training; 3 days of extinction training at 9 trials per day were employed; *S* was allowed 90 sec. to traverse the runway in extinction; there was no extinction criterion, i.e., each *S* was given all 27 extinction trials; a consistent group was not employed; and, of course, the intertrial interval was longer here, it never being less than 20 min. either in acquisition or extinction. Group SA ($N = 11$) was given 3 days of acquisition training according to a single alternating pattern of reinforcement, 9 trials per day, the initial daily trial being reinforced. Group R was rewarded according to the same schedule employed by Capaldi and Hart (1962). Thus, for Group R there was only a single daily transition from a nonreinforced to a reinforced trial.

Results and discussion.—The performance of the groups on each day of acquisition and extinction is shown in Fig. 1. Group SA ran slightly but not significantly faster than Group R on Days 2 and 3 of acquisition, a result consistent with that obtained by Capaldi and Hart (1962), for Day 2, $F(1, 20) = 1.24$; for Day 3, $F(1, 20) = 1.72$, using log times as the score. On each day of extinction Group R ran slower than Group SA, a result which is also consistent with that obtained under massed-trial conditions. The log times in extinction were summed and the resulting analysis yielded an $F(1, 20) = 12.95$ which is significant well beyond the .01 level.

It is considered that the present results can be taken to indicate that aftereffects remain functional for at least 20 min., a period which exceeds the modest upper limit of 15 min. suggested by Sheffield (1949). Aftereffects are not observable as such; therefore, any specification of their properties, temporal or otherwise, will necessarily have to be inferential. Accordingly, it can be argued that if a prediction entailed by Hull's

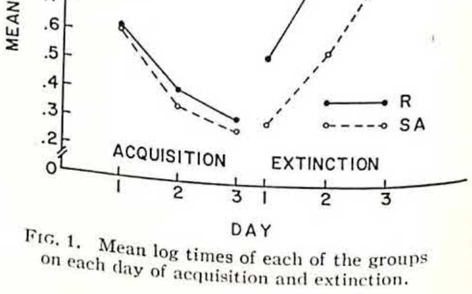


FIG. 1. Mean log times of each of the groups on each day of acquisition and extinction.

(1952) hypothesis can be confirmed under conditions of relatively spaced trials, then a basis exists for assuming that aftereffects remain functional for at least the duration of the temporal interval employed. Capaldi and Hart (1962) have shown in some detail that the present results follow directly from Hull's hypothesis; hence the conclusion suggested above that aftereffects can be considered to remain functional for at least 20 min.

REFERENCES

CAPALDI, E. J., & HART, D. Influence of a small number of partial reinforcement training trials on resistance to extinction. *J. exp. Psychol.*, 1962, **64**, 166–171.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

(Received April 18, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 3, 319–320

KINESTHETIC AFTEREFFECT AND MODE OF EXPOSURE TO THE INSPECTION STIMULUS[1]

PAUL BAKAN AND ERNEST WEILER

Michigan State University

Measures of kinesthetic aftereffect were taken on 3 groups of 48 Ss, differing in type of exposure to an inspection stimulus. Largest aftereffects were found for the group experiencing passive movement and smallest aftereffects were found for the group experiencing no movement of the inspection stimulus between thumb and index finger. A group experiencing active movement showed aftereffects intermediate in size between the other 2 groups. Results support a hypothesis of a relationship between attention, neural satiation, and aftereffects.

Bakan and Thompson (1963) have studied the effect of two modes of exposure to the inspection stimulus (I stimulus) on the size of the kinesthetic aftereffect. These are a Move condition where *S* moves a hand back and forth along the I stimulus, and a Hold condition where *S* merely holds the I stimulus. On the basis of an extension of satiation theory which postulates a relationship between degree of attention, amount of satiation, and size of aftereffect (Köhler & Adams, 1958) it was predicted that the Move condition would produce a larger aftereffect because the movement would enhance attention directed to the I stimulus. The difference between the Move and Hold condition, while not statistically significant, was in the expected direction.

In retrospect, it appears that this experiment did not represent a fair test of the attention-satiation hypothesis. The Move condition involved the use of major arm muscles which by contributing to the total afferent kinesthetic stimulation may have served to attenuate the effect of the relevant kinesthetic stimulation from muscles in the hand. These large arm movements were absent in the Hold condition. The present experiment was designed to control for this extraneous variable by the introduction of a Passive Move condition in which the I stimulus is moved back and forth between *S*'s thumb and finger while the arm is kept stationary.

Method.—There were 48 undergraduate *S*s in each of three groups, Active Move (AM), Passive Move (PM), and Hold (H). The apparatus consisted of three wooden blocks: (*a*) a standard block 1.5 in. wide, 5 ft. long, and 2 in. thick, (*b*) a variable-width tapered block 60 in. long and 2 in. thick ranging in width from 1 in. to $2\frac{5}{16}$ in. and changing at the rate of $\frac{1}{32}$ in. per inch, and (*c*) an I stimulus 0.5 in. wide, 2 ft. long, and 2 in. thick which could remain stationary while *S* held it, or while he moved the hand along it,

---

[1] This work was supported by a grant from the National Science Foundation. This paper was written during the senior author's stay at the Medical Research Council, Applied Psychology Research Unit in Cambridge, England, 1961–62.

The authors wish to thank Joseph Levine for assistance in planning this experiment.

or could be moved back and forth between the thumb and forefinger by *E*.

The blocks were mounted on two tables 34 in. high and parallel to each other. The standard and I block were at *S*'s left and the variable at his right. The variable block was fitted with an H shaped rider within which *S* placed his thumb and forefinger while making equality matches.

The *S*, while blindfolded held the standard between the thumb and forefinger of the left hand, while with the right hand he found a point on the variable block which felt equal in width to the standard. A series of four practice judgments was followed by four preinspection judgments. The four judgments were in an ABBA order from points 1.25 in. and 1.75 in. in width. The mean of four judgments constituted the point of subjective equality (PSE) used in analysis.

The inspection period (I period) followed determination of the preinspection PSE and was followed by four judgments whose mean constituted a postinspection PSE. The difference between the pre- and postinspection PSEs constitutes the measure of the aftereffect. In the I period *S*s in the H group held the I stimulus between the thumb and index finger for 60 sec.; *S*s in the AM group held the I stimulus between thumb and index finger while moving the hand back and forth along it in about 90 excursions of about 18 in.; *S*s in the PM group held the I stimulus between thumb and index finger with hand stationary while *E* moved the I stimulus back and forth over approximately 90 excursions of about 18 in.

Results and conclusions.—The mean aftereffect for the H, AM, and PM groups, respectively, was .068, .085, and .120 in. The between-group variance is significant at the .05 level. Further tests show that the difference between the H and PM conditions is significant at the .01 level; the difference between the AM and PM conditions falls just short of significance at the .05 level; and the difference between the H and AM conditions is not statistically significant.

As in the earlier study the difference between the H and the AM condition, though in the direction predicted by the attention-satiation hypothesis, was not statistically significant. The elimination of gross muscle movements in the PM condition resulted in a significantly larger aftereffect than that produced by the H condition. In view of the rationale for the introduction of the PM condition cited earlier, the comparison between the H and PM conditions appears to constitute a fair test of the attention-satiation hypothesis. The results support the postulated relationship between attention and satiation.

REFERENCES

BAKAN, P., & THOMPSON, R. Attention and kinesthetic after-effects. *Amer. J. Psychol.*, 1963, **76**, in press.

KÖHLER, W., & ADAMS, P. Perception and attention. *Amer. J. Psychol.*, 1958, **71**, 489–503.

(Received May 3, 1962)

Journal of
Experimental Psychology

Vol. 65, No. 4 April 1963

EFFECT OF INSTRUCTIONS AND INFORMATION ON PROBLEM SOLVING[1]

CARL P. DUNCAN

Northwestern University

In solving problems which involved turning switches to produce certain patterns among lights, male Ss who were strongly instructed to minimize overt responses (switch turns) and to think, made fewer overt responses and took, usually, longer time to solve than uninstructed Ss. These effects of instructions occurred with each of 3 different amounts of information (levels of difficulty) about a problem, and on an initial as well as on a transfer problem. Mild instructions to minimize overt responses, given to other groups, had no effect. Uninstructed female Ss solved problems as well as uninstructed male Ss, but performance of female Ss was not significantly affected by any of the instructions.

As a variable in problem solving, verbal instructions to *S* have often been shown to influence performance on simple set problems such as anagrams or water jars (Duncan, 1959). But with more complex, particularly insight-type problems, instructions or lectures have had little or no effect. Maier (1933) reported that a higher percentage of *S*s who had been given a lecture on problem solving solved three insight problems than a nonlectured control group. However, the lectured group was actually better on only one of the three problems, and it was reported that not all of the control *S*s had time to work on two of the problems. In two attemps to replicate, in part, Maier's study, using several different kinds of instructions or a short lecture, there was no difference between instructed and uninstructed groups in solving the two-string problem with pendulum solution (Duncan, 1961).

When *S* is attempting to solve an insight problem, there may be little or no overt behavior that can be detected, particularly if the problem is presented in written form. Since *S*'s responses to such problems are almost entirely implicit (thinking), it is difficult to determine if instructions influenced *S* in any manner. In the present study a different kind of fairly complex problem is used, one in which *S* must make some overt responses in order to discover the solution. Problems of this kind, in which the solution consists of some combination or pattern among buttons to be pushed, switches to be turned, etc., not only demand a certain amount of overt responding by *S*, but also often permit some aspects of the problem to be discovered by thinking, e.g., inferring hidden connections between lights and switches (French & Thomas, 1958; John, 1957).

---

[1] This study was supported by Grant M-3182 from the Institute of Mental Health, National Institutes of Health, United States Public Health Service. Thanks are due Keith Clayton, Peter Delaney, David Isaacs, Suzanne Meyer, and Robert Rehula.

In the present study, the basic comparison is between uninstructed *S*s who are allowed to solve a problem in their own way, using whatever combination of overt and implicit activity they choose, and *S*s who are instructed to minimize overt behavior and to try to arrive at a solution as much as possible by thinking. The interest is in determining what effects nonspecific instructions to think have on *S*'s problem solving behavior. Two experiments are reported, differing primarily in the extent to which *S* is given information about the problem to be solved.

EXPERIMENT I

Method

Apparatus.—The apparatus was a copy of one designed and used by E. G. French (French, 1958; French & Thomas, 1958). The sloping panel front of a metal cabinet contained a horizontal row of seven lights, with a two-position toggle switch below each light. A strip of cardboard, carrying No. 1–7, was fixed in the space between lights and switches. The wiring is such that each light is controlled by more than one switch, and each switch controls more than one light. In addition, turning a switch *either* on or off always reverses the state, from on to off or off to on, of any light to which that switch is connected. Since any number of lights from 0 to 7 may be either on or off, there are 2^7 or 128, possible patterns of lights, each of which is produced by one, and only one, of the 128 possible combinations of switches on or off. The particular switches that must be turned on or off to produce a particular pattern of lights may be turned in any order.

In the present study a problem is defined as a certain pattern of lights, specified verbally by *E* for *S*, which *S* must produce by turning switches. Two problems were used. Problem 1 for all *S*s was to get Light 2 on, all other lights off, the solution to which required turning on six of the seven switches (No. 1, 2, 3, 4, 5, and 7). Problem 2 for all *S*s was to get all lights on except Light 2; the solution was Switches 2, 5, and 6 on. When *S* begins to solve a problem, all lights are off. Solution of a problem is achieved by turning switches on or off until the pattern of lights specified by *E* appears.

The measures taken were total number of switch turns (responses), and total time, to solve a problem. Each turn of a switch, either on or off, was counted by a counter sitting beside the apparatus in plain sight of *S*. The *E* used a stop clock to measure total time from the moment when *E* completed his instructions to *S* until *S* solved the problem.

Subjects.—The *S*s were 264 students from introductory psychology courses, 44 men and 44 women in each of the three groups. The *S*s were assigned to groups in turn.

Conditions.—There were three conditions in Exp. I with an independent group of *S*s for each condition. The conditions were differentiated only by the kind of instructions given prior to Problem 1. In Group 1 (uninstructed) the instructions were kept to a minimum; after an *S* in this group had sat down at the table on which the apparatus was placed, *E* said:

> You see that this apparatus has a set of lights and a set of switches. Here is what I want you to do. Using these switches, turn on Light 2 only, that is, Light 2 on, all other lights off. One more thing. Turn each switch on or off *individually*; do not turn several switches on or off simultaneously. Of course, you can have as many switches on or off at any one time as you like. Use only one hand.

For Group 2 (mild instructions), the first three sentences of the instructions were the same as those given above for Group 1. Then *E* also said, "This counter (*E* points to it) counts every turn of a switch either on or off. The more it counts, the worse your score is. One more thing . . ." (the rest of the instructions were the same as for Group 1).

For Group 3 (strong instructions), the first three sentences of the instructions were the same as for Groups 1 and 2. Then *E* continued:

> Now before you start, let me point out something about this problem. You can't

solve the problem without pushing some switches, but you can solve the problem by pushing very few switches *if* you do some *thinking*. Don't just start flipping switches on and off to see what happens; a monkey can do that (slight pause). Now, this counter counts every turn of a switch either on or off; the more it counts, the worse your score is (slight pause). One more thing. Turn each switch on or off *individually*; do not turn several switches on or off simultaneously. Of course, you can have as many switches on or off at any one time as you like. Use only one hand. O.K., go ahead, and remember to *think*. If you use your head you can solve the problem of getting only Light 2 on by turning very few switches.

After an *S* in any group had solved Problem 1, *E* recorded total switch turns and total time, reset the clock to zero, and turned off all switches (taking less than 10 sec.). Then *S* was started on Problem 2 by *E*'s saying, "O.K., now get all lights on *except* Light 2, that is, Light 2 is to be off, all other six lights on. Go ahead." These were the only instructions given to any *S* immediately prior to Problem 2.

Any *S* who had not solved Problem 1 in 20 min., or both problems in 30 min., was rejected and replaced by another *S*. One male and 1 female *S* in Group 1, 1 male *S* in Group 2, and 10 male and 7 female *S*s in Group 3 were rejected for failure to meet these criteria. Thirteen of the 17 *S*s rejected in Group 3 were apparently too severely inhibited by the instructions; they made too few overt responses. These 13 *S*s averaged less than one switch turn per 10 sec.; very few *S*s with this slow a rate of overt response managed to solve even Problem 1. It is possible that the slower *S*s eliminated from Group 3 were merely poor problem solvers. But it also is possible that if these *S*s had been allowed to continue to work on the problem, their scores for number of switch turns would have been among the lowest, since, as will be seen, the instructions given Group 3 tended both to increase solution time and decrease number of switch turns (thus slowing rate of response).

Results

Problem 1.—The top portion of Table 1 shows mean responses (switch turns) to solution for men and women separately in the three groups. Analysis of variance of mean responses for men yielded F (2, 129) = 4.05, $p < .02$. (Because of positive skewness, the scores were also analyzed after conversion to logarithms; for log scores, $F = 4.33$.) By t test, Groups 1 and 2 did not differ ($t < 1$); Group 3 differed significantly both from Group 1 ($t = 2.63$) and from Group 2 ($t = 2.27$). Thus, in comparison to the uninstructed *S*s (Group 1), the mild instructions given Group 2 were ineffective, but the strong ininstructions given Group 3 significantly reduced number of overt responses to solution.

TABLE 1

MEANS FOR NUMBER OF RESPONSES AND FOR TIME ON PROBLEM 1 IN EXP. I

Group	Number of Responses			
	Men		Women	
	Mean	σ_m	Mean	σ_m
1	143.57	18.36	142.32	15.80
2	134.91	21.05	134.93	20.83
3	80.07	9.91	95.34	10.87
	Time in Sec.			
1	229.73	32.10	258.93	28.66
2	246.50	29.13	274.86	38.27
3	365.89	44.39	347.00	40.19

Mean responses for women showed the same trend among groups as the means for men, but the differences among women's means were not significant (F for raw scores was 2.39, for log scores, 1.28). Table 1 shows that in Groups 1 and 2, men and women were not different, a fairly rare finding in problem solving. In Group 3, instructions affected women in the same direction as men, but not significantly so.

The bottom portion of Table 1 shows mean time in seconds to solution. It can be seen that the trend among mean times is in the opposite direction as that for mean responses. Analysis of men's means for time gave F (2, 129) = 4.30, $p < .02$. (F for log sec. was 3.56, $p < .05$). By t test, Groups 1 and 2 did not differ ($t < 1$); Group 3 was significantly slower than either Group 1 ($t = 2.69$) or Group 2 ($t = 2.36$). Analysis of women's means gave $F = 1.69$, not significant. Thus, although in the opposite direction, the results for time scores are essentially the same as for responses; instructions significantly influenced performance in Group 3 men, but not in women.

Problem 2.—Table 2 shows means for responses and for time for all groups on Problem 2. Among men, the rank order of group means, for both measures, is the same as on Problem 1, but the means for women are irregular. Analysis of mean responses for men gave F (2, 129) = 3.43, $p < .05$. Again Groups 1 and 2 did not differ, but Group 3 made significantly fewer responses than Group 1 ($t = 2.49$), and nearly so in comparison to Group 2 ($t = 1.95$). Analysis of mean responses for women yielded $F = 1.77$, not significant. The means for time do not differ; $F = 1.11$ for men, 2.43 for women.

TABLE 2

MEANS FOR NUMBER OF RESPONSES AND FOR TIME ON PROBLEM 2 IN EXP. I

Group	Number of Responses			
	Men		Women	
	Mean	σ_m	Mean	σ_m
1	91.55	13.34	74.09	11.98
2	83.20	10.77	87.82	17.98
3	52.86	8.31	74.27	12.21
	Time in Sec.			
1	144.84	24.74	128.00	23.12
2	151.25	20.24	185.05	38.41
3	197.98	35.23	234.82	39.01

Thus, without further instructions to think, Group 3 men continued to perform on Problem 2 as they had on Problem 1, solving the problem with fewer overt responses than either of the other two groups. And on Problem 2, Group 3 men still took longer, though not significantly longer, to solve.

EXPERIMENT II

In Exp. I, *S*s had been given no information about the problem, about the connections between lights and switches. The purpose of Exp. II was to determine if instructions would influence performance when *S*s had more information about the problem device.

Method

Apparatus.—This was the same as in Exp. I except that a horizontal strip of white cardboard (cue card) was fixed to the apparatus between the row of lights and the row of switches. (The strip carrying the numbers 1–7 was placed above the lights.) Arrows drawn on the cue card ran from each switch to each of the lights to which that switch was connected.

Conditions.—A 2 × 2 design was used. Groups were either uninstructed or instructed, and were given either "some" information or "complete" information about the operation of the problem apparatus. Uninstructed groups were given the same minimal instructions as was Group 1 in Exp. I. Instructed groups were given the same "strong" instructions as was Group 3 in Exp. I.

As in Exp. I, after *S* had sat down facing the apparatus, *E* first said, "You see this apparatus has a set of lights and a set of switches." Then for groups given Some Information, *E* continued:

> See these arrows (pointing to the arrows on the card)? They show what switches are connected to what lights. An arrow leading from a switch to a light means that that switch is connected to that light. You can see that each switch is connected to more than one light.

The *E* then concluded with instructions appropriate for either Uninstructed or Instructed conditions (as in Exp. I).

For groups given Complete Information, the initial portion of the instructions was the same as that given above for Some Information. Then *E* continued:

> Now I am going to tell you what these connections mean; listen carefully and remember. If a light is *off*, turning a switch that is connected to that light will always turn the light *on*. If a light is already *on*, turning a switch that is connected to it will always turn the light *off*. Do you understand? Turning a switch either up or down always reverses the state, from on to off, or off to on, of all lights that are connected to that switch. Is that clear?

At this point *E* tried to make sure *S* understood. The *E* then concluded with the instructions appropriate for either Uninstructed or Instructed conditions.

Other details of procedure were the same as in Exp. I. After *S*s had solved Problem 1, they were given Problem 2 without further instructions. The cue card remained on the apparatus for both problems.

Subjects.—The *S*s were male students from introductory psychology courses, assigned in turn to the four groups: Group US (uninstructed-some information), IS (instructed-some information), UC (uninstructed-complete information), and IC (instructed-complete information). There were 23 *S*s in each of the four groups.

Results

Problem 1.—The top half of Table 3 shows mean responses to solution for the four groups. A 2 × 2 analysis of variance of these means yielded F (1, 88) = 11.12, $p < .0025$, for the Instruction variable, F (1, 88) = 16.44, $p < .001$ for the Information variable. The interaction was not significant, F (1, 88) = 1.07. (Analysis of log scores gave $F = 13.96$ for Instruction, 18.96 for Information, <1 for the interaction.) The overall mean for the two uninstructed groups combined ($N = 46$) was 46.98; for instructed groups combined, the mean was 22.78. For some information, the overall mean was 49.59; for complete information, 20.17. Thus, as compared to Exp. I (Table 1), where *S*s were given no information, overt responses to solution were fewer when *S*s were given some information, and considerably fewer when they were given complete information, about the problem device. Within each of these degrees of information, instructions to think materially reduced the number of overt responses to solution.

TABLE 3

MEANS FOR NUMBER OF RESPONSES AND FOR TIME ON BOTH PROBLEMS IN EXP. II

Group	No. of Responses		Time in Sec.	
	Mean	σ_m	Mean	σ_m
Problem 1				
US	65.43	11.94	246.17	33.64
IS	33.74	7.19	328.78	43.40
UC	28.52	6.08	175.17	34.68
IC	11.83	2.52	218.09	34.84
Problem 2				
US	27.09	6.54	107.96	38.04
IS	19.17	5.65	151.52	38.13
UC	26.48	5.71	88.87	5.34
IC	10.56	2.34	135.78	36.57

Mean time to solution, also shown in Table 3, again indicates some tendency for longer times to be associated with conditions in which *S*s were instructed to think. However, analysis of the means revealed that F for the Instruction variable was 2.90, not significant. For the Information variable, F (1, 88) = 6.08, $p < .02$. The interaction term gave $F < 1$. Overall means (in seconds) were 210.67 for uninstructed, 273.43 for instructed groups, and 287.48 for some information, 196.63 for complete information groups. Thus, those *S*s given some information took longer to solve, probably because they made a greater number of overt responses than *S*s given complete information. Instructed *S*s had (numerically) longer times, along with fewer overt responses, than uninstructed *S*s.

Problem 2.—Mean responses are shown in the bottom half of Table 3. Analysis indicated that the Instruction variable continued to be effective, F (1, 88) = 5.03, $p < .03$, but the Information variable was not ($F < 1$). The interaction term also gave $F < 1$. Overall means were 26.78 for uninstructed, 14.87 for instructed groups, and 23.13 for some information, 18.52 for complete information groups.

The means for time, also given in Table 3, show much the same pattern as for Problem 1, i.e., longer times for instruction, and for some information conditions. However, there were no significant differences among time scores, $F = 1.81$ for instruction, <1 both for information and for the interaction.

DISCUSSION

The purpose of the study was to attempt to influence *S*'s problem solving behavior by instructions to think. As mentioned earlier, previous attempts to affect any aspects of *S*'s performance by instruction did not appear to have been particularly successful with problems more complex than simple set tasks (Duncan, 1959, 1961; Maier, 1933). In the present study the results of both experiments indicate that with "strong" instructions, with male *S*s, and with the kind of problem used, problem solving behavior was modified in significant ways. Specifically, the male *S*s who were strongly instructed to minimize overt responses and to think, arrived at solution with significantly fewer overt responses while taking (usually) longer times to solve. It was also shown that these two effects of the instructions carried over to a second problem.

It may be of interest to examine the quantitative extent to which strong instructions influenced the two measures, responses and time, at each of the three degrees of information; i.e., the "no information" of Exp. I, and the some and complete information conditions of Exp. II. If one compares, on Problem 1, Group 1 with Group 3 men in Exp. I (Table 1), and uninstructed with instructed groups under both information conditions of Exp. II (Table 3), the following trends appear. Both for responses and for time, the absolute difference between uninstructed and instructed means decreases as one goes from no to some to complete information. But since different amounts of information produced sharp changes in the level of difficulty of the problem, the relative differences (percentages computed by taking the absolute difference between instructed and uninstructed means and dividing by the uninstructed mean) seem more appropriate. For this measure, the trends for responses and for time are in opposite directions. The percentage *decrease* in mean responses produced by instructions was 44.2%, 48.4%, and 58.5% for no, some, and complete information, respectively. Thus there was some increase in the relative effect of instructions on responses, though it was not great (and in Exp. II it was shown that the interaction between instruction and information was not significant). For time, the percentage *increase* in mean attributable to instructions was 59.3%, 33.5%, and 24.5% for the same three degrees of information. So there was a fairly sharp decrease in the effect of instructions on time scores as a function of information.

Since instructions (hereafter, the "strong" instructions given Group 3 in Exp. I and instructed groups in Exp. II) reduced the number of overt responses, in men, prior to solution, it seems reasonable to infer that the instructed men did attempt to carry out some portion of their problem solving by whatever implicit activity they considered to be thinking. But since these *S*s took longer to arrive at solution, it might be argued that their thinking, or more specifically their combination of overt and implicit activity, was inefficient because it was too time consuming. Although taking time to think might have disadvantages in some cases, usually it is only the possible advantages that are mentioned. Dollard and Miller (1950) list "training to stop and think" (p. 118) as part of social training of the higher mental processes. Their argument is that if one makes an immediate instrumental response to a problem, there may not be enough time for appropriate cue-producing responses to occur. Highland, Newman, and Waller (1956), in describing differences between successful and less successful mechanics in trouble-shooting, suggested that a mechanic should not be too hasty in starting to make checks (overt responses) with his test instruments, that skill in locating malfunctions would be improved if mechanics were taught to exercise restraint in making checks, especially at the beginning of the problem. Their description of the more successful trouble-shooters strongly suggests that this group takes time to think about the overt responses they do make.

In a study of pattern conception, Shipstone (1960) instructed *S*s in some conditions to slow down, in other conditions to slow down and to use logical procedures while categorizing. For some classes of stimuli, these instructions improved performance, largely, it was concluded, because the instructions slowed *S*s down and made them "stop, look, and think more carefully" (p. 24). Also, longer time scores were correlated with more successful categorizing strategies.

Thus, some problems, such as those used in the present study, those used by Shipstone, and some others (Duncan, 1959) apparently can be solved either by largely overt trial and error behavior, or through partly overt and partly implicit behavior. (This is in contrast to the typical insight problem which, particularly when in pencil and paper form, forces *S* to rely almost entirely on implicit activity.) With such problems, instructions, and perhaps other variables, can produce variation in the degree to which *S* relies on overt response, and so, presumably, in the degree to which he engages in some kind of implicit activity, or thinking. As the ratio of overt to implicit activity decreases, there may be an increase in time spent on the problem. These considerations suggest that for a particular problem, the optimum ratio of overt to implicit activity for one criterion, e.g., learning how to categorize, may not be the optimum ratio for some other criterion, such as solution of the problem.

REFERENCES

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

DUNCAN, C. P. Attempts to influence performance on an insight problem. *Psychol. Rep.*, 1961, **9**, 35–42.

FRENCH, E. G. The interaction of motivation and ability in problem-solving success. *J. abnorm. soc. Psychol.*, 1958, **57**, 306–309.

FRENCH, E. G., & THOMAS, F. H. The relation of achievement motivation to problem-solving effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45–48.

HIGHLAND, R. W., NEWMAN, S. E., & WALLER, H. S. A descriptive study of electronic trouble shooting. In G. Finch & F. Cameron (Eds.), *Air force human engineering, personnel, and training research.* (NAS-NRC Publ. No. 455) Washington, D. C.: National Academy of Sciences-National Research Council, 1956.

JOHN, E. R. Contributions to the study of the problem-solving process. *Psychol. Monogr.*, 1957, **71**(18, Whole No. 447).

MAIER, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, **24**, 144–155.

SHIPSTONE, E. I. Some variables affecting pattern conception. *Psychol. Monogr.*, 1960, **74**(17, Whole No. 504).

(Received May 1, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 328–333

SELECTIVE ATTENTION AND CODING IN VISUAL PERCEPTION[1]

CHARLES S. HARRIS
Harvard University

AND

RALPH NORMAN HABER
Yale University

The form of S's silent verbal encodings of briefly presented stimuli was manipulated. Order of encoding accounted for the effects of instructions to attend selectively on accuracy of report. Instructions to attend selectively had no significant effect on Ss who encoded the stimulus in the order dictated by English syntax. The Ss whose coding strategy permitted them to encode first the stimulus attribute they were told to attend to reported this attribute more accurately than incidental attributes. The superior accuracy was not a by-product of order of report. The attribute encoded 1st is based on a better visual trace and also may be retained better in short-term memory.

Attending to a single property of an object allegedly results in clearer and more vivid perception of that property (Bruner, 1957; Chapman, 1932; Humphrey, 1951; Külpe, 1904). Perception of "incidental" properties is supposedly inhibited or "gated out."

When a stimulus is presented briefly, *S*s report incidental aspects less accurately than aspects they were told to attend to (Chapman, 1932; Külpe, 1904; Lawrence & LaBerge, 1956; Wilcocks, 1925). But if *S*s usually report the attended-to aspect first, inaccurate reporting of incidental aspects may be due simply to forgetting, rather than to faulty perception (Rubin, 1913).

Lawrence and LaBerge (1956) found that the difference in accuracy between first and later reports (when *S* was told to pay equal attention to all attributes of the stimulus) was about the same as that between "crucial" and incidental aspects when *S* was instructed to attend selectively. They concluded that such instructions do not affect initial perception, but only the order in which *S* reports the attributes.

However, Wilcocks (1925) asserted that when *S*s were told, immediately after the stimulus was flashed, which one of three attributes to report, they reported incidental attributes less accurately than attributes they attended to. Although only some of Wilcocks' supporting data show statistically reliable differences, the possibility remains that attention does affect initial perception.

In an unpublished pilot study[2] by the present investigators, only 2 out of 77 *S*s relied on a visual image when reporting a briefly presented stimulus. The remaining 75 encoded the stimulus into verbal form and silently rehearsed the verbal encoding until the time came to report it. Inadequate encoding of incidental aspects of the stimulus might account for the difficulty in reporting them.

---

[1] This research was supported in part by a grant (MHO 3244) from the United States Public Health Service to the second author, and was conducted during the senior author's tenure as a National Science Foundation Graduate Fellow. We wish to thank Jeremy Nahum, Donald Richard, and J. Michael Ross for their help in gathering data.

[2] Individual *S*s' means, statistical analyses, and data on Free order of report, as well as further discussion, information on pilot studies, and statistical tests of Wilcocks' data, have been deposited with the American Documentation Institute. Order Document No. 7451 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Miller (1956) has stressed the role of coding in memory, and other investigators have found that coding affects some measures of visual perception. For instance, the labels attached to a stimulus may influence drawings of it (Carmichael, Hogan, & Walter, 1932; Herman, Lawless, & Marshall, 1957; Kuhlman, 1906; Moore, 1910). Recently, Sperling (1963) has included verbal encoding in his model for visual memory tasks. He notes that many errors in identifying briefly exposed letters seem to be auditory rather than visual confusions—e.g., confusions between D and T rather than between E and F (Sperling, 1960, p. 21).

Thus, the effect of instructions to attend selectively has been attributed to an effect on perception, order of report, or coding. In order to determine the cause of the selective effect, the present investigators manipulated the form of *S*'s code, order of report, and instructions about attention.

METHOD

Subjects

Forty-eight unpaid Yale University undergraduates served, 1 to 4 at a time, as *S*s.

Stimuli

Each stimulus, like those used by Lawrence and LaBerge (1956), consisted of a pair of cards from the Wisconsin Card Sorting Test (available from Kilgore Printing Company, Madison, Wisconsin). The cards differ in the number of figures portrayed (one, two, three, or four), their shapes (stars, circles, crosses, or triangles), and their colors (red, blue, green, or yellow). Eighty pairs of cards were selected randomly, with the restriction that the two cards differ on each attribute.

Codes

Before the experimental session began, each *S* was told to use one of the two basic types of code that *S*s in the pilot study used spontaneously: *Objects code*, which describes the stimulus as two sets of objects (e.g., "two red circles, four blue stars"), or *Dimensions code*, which describes the three dimensions of the stimulus (e.g., "red, blue; two, four; star, circle").

To encourage Objects coding, a single card (half of the usual stimulus) was exposed for $\frac{1}{10}$ sec., *S* wrote down what he had seen, and *E* gave the correct answer in Objects code. This was repeated with 11 more single cards. Then *E* said: "We have found that you will do best if, as soon as the flash comes, you repeat to yourself what you saw, . . . for example, two blue triangles, three green circles." The *S* then practiced on six two-card stimuli.

To encourage Dimensions coding, 12 pairs of cards were exposed, but *S* was told to write down only one dimension—e.g., colors—of each stimulus. The *E* gave the correct answer—e.g., "blue, green"—for only that dimension. Then *E* made the same statement as above, changing the ending to: ". . . for example, two-three, blue-green, triangle-circle." Six more pairs of cards were presented, with *S* writing down all three dimensions.

At the end of the experimental session, *S* answered an extensive questionnaire about his coding strategy. Two judges (who did not know which *S*s had been trained to use Dimensions code and which to use Objects code) examined the answers to determine which of four codes each *S* had actually used: *Objects code* or *Dimensions code*, as defined above; *Modified Objects code*, which is an Objects code changed in various ways to stress the crucial dimension (e.g., "red, blue; two red circles, four blue stars" or "two red circles, four blue stars; red, blue"); or *Incomplete Dimensions code*, which is a Dimensions code that encodes only part of the stimulus (e.g., "red, blue; two, four; . . .").

The six disagreements between the judges were resolved by jointly re-examining the answers to the questionnaire (in five instances, the final decision was that the *S* did not consistently use a single code).

Instructions about Attention

Before each stimulus was flashed, *S* read instructions about to what to attend. There were three types of instruction:

Emphasis instruction.—The *S* was told to pay special attention to one of the three aspects of the stimuli (*Emphasized dimension*), since he would receive 30 points for reporting that aspect correctly, but only 3 points for each of the other two aspects (*Unemphasized dimensions*). The *S* was told that he would not be given any money for the points, but that "the points are awarded as a convenient index of value, and therefore we hope that you will feel motivated to try to get as many as possible. . . . I will be happy to tell you how well you did . . . after we have analyzed the results."

Equal instructions.—The *S* was told to pay equal attention to all three aspects, as he would receive 15 points for each one correctly reported.

One-Only instructions.—The *S* was told to pay attention to and report only one dimension, for which he would receive 30 points if correct. He would receive no points for reporting the other two dimensions.

Presentation of Stimuli

Each *S* reported on 80 stimuli: 48 with Emphasis instructions (16 each with colors worth 30 points, numbers worth 30 points, and shapes worth 30 points), 16 with Equal instructions, and 16 with One-Only instructions. The order of stimuli was the same for all *S*s, but the instruction sheets and orders of report varied from *S* to *S* for any given stimulus.

Each stimulus was flashed on a screen for $\frac{1}{10}$ sec., after which *S* had 15 to 20 sec. to record what he had seen. The stimulus on the screen measured 40 × 15 in. (approximately 10° × 4° of visual angle). There was no fixation point. The room was illuminated just brightly enough to enable *S* to read the instructions and check lists without difficulty. Except during the stimulus flash, the screen was illuminated only by the dim room lights.

Order of Report

Order of report was not known to *S* until after each stimulus was presented. On half of the trials (*Forced order of report*), *S* turned over the instruction page immediately after the stimulus was flashed and found a check list for one of the three dimensions. He put check marks next to two values of that dimension. Then he turned to the next page, checked off the second dimension, and so on.

On the remainder of the trials (*Free order of report*), *S* found a page with check lists for all three dimensions and could report them in any order (except that he had to check both values of each dimension before proceeding to the next). After recording all three dimensions, he numbered them to indicate his order of report.

RESULTS

The attempt to manipulate *S*'s coding strategy was quite successful, as Table 1 shows. The results reported below do not include data from the 10 *S*s who did not consistently use a single code and the 5 *S*s who did not use the code they were trained to use.

Unless otherwise noted, all statements refer to performance with Forced order of report. In general, however, the same statements apply to performance with Free order of report. Analyses of variance showed that freedom of order of report produced neither a significant main effect nor significant interactions with attention instructions, position of stimulus, or order of report.

Main Finding

The main finding was that the code *S* used determined whether or not his accuracy was affected selectively by attention instructions. The interaction between codes and instructions is significant beyond the .01 level.

Three of the four coding groups made more errors in reporting Unemphasized dimensions than in reporting Emphasized dimensions (Table 2). The difference between the mean percentage of errors on Emphasized

TABLE 1

NUMBER OF *S*s WHO WERE TRAINED TO USE AND WHO REPORTED USING EACH CODE

Code That *S* Reported Using	Code That *S* Was Trained to Use	
	Objects	Dimensions
Objects	8	3[a]
Modified Objects	9	1[a]
Dimensions	1[a]	9
Incomplete Dimensions	0	7
Inconsistent	5[a]	5[a]

[a] Data discarded.

TABLE 2

Mean Percentage of Errors with Forced Order of Report

Code	Emphasis Instructions						Equal Instructions			One-Only Instructions
	Emphasized Dimension			Unemphasized Dimensions						
	First	Second	Third	First	Second	Third	First	Second	Third	
Objects	10.4	9.0	8.3	7.7	7.1	9.2	7.0	9.4	14.8	4.1
Dimensions	3.5	8.0	10.5	12.1	15.7	15.0	2.8	4.2	9.0	3.4
Incomplete Dimensions	8.3	7.9	11.7	18.3	22.8	35.0	12.0	21.7	23.0	5.9
Modified Objects	5.4	11.4	9.9	16.0	23.3	18.7	14.7	15.6	18.0	5.8

Note.—Column entries refer to errors made when the specified dimension was reported first, second, or third. Each *S* had 48 opportunities for error on the Emphasized dimension, 96 on the Unemphasized dimensions, 48 under Equal instructions, and 32 under One-Only instructions.

and on Unemphasized dimensions is significant beyond the .02 level for Dimensions coders ($t = 2.96$), beyond the .01 level for Incomplete Dimensions coders ($t = 4.10$), and beyond the .02 level for Modified Objects coders ($t = 2.93$).

The higher rate of error on Unemphasized dimensions was evident even on first reports, indicating that it is not a by-product of order of report. Considering first reports only, the difference between rates of errors on Emphasized and on Unemphasized dimensions is significant beyond the .01, .02, and .02 levels ($t = 3.55$, 3.54, 3.08) for the Dimensions, Incomplete Dimensions, and Modified Objects coders, respectively. Hence, "the act of writing down the first dimension plus the increased temporal delay" (Lawrence & LaBerge, 1956, p. 17) cannot be responsible for the less accurate reporting of Unemphasized dimensions.

How, then, does this relative inaccuracy come about? We hypothesize that *S*s wanted to give the "more valuable" dimension preferential treatment. The questionnaire revealed that 7 of the 9 Dimensions coders, all 7 of the Incomplete Dimensions coders, and 3 of the 9 Modified Objects coders always encoded the Emphasized dimension first (the other 6 Modified Objects coders did so frequently).

An item encoded first gains a triple advantage. (*a*) It is encoded more accurately, from a detailed visual image that fades markedly before *S* can encode the remaining aspects (Sperling, 1960). (*b*) It may be better retained during rehearsal (studies on immediate memory have seldom eliminated the effects of order of encoding, and provide no clear-cut answer about which part of a list is retained best; see, for example, Deese & Kaufman, 1957; Kay & Poulton, 1951; Pollack, 1952). (*c*) It is conspicuously placed, so that it can be reported first, when the opportunity arises, and quickly, without running through the entire encoded list.

An Objects coder, unlike *S*s using the other codes, cannot easily give special treatment to the Emphasized dimension. To do so he would have to go against long-standing habits based on standard English word order. And indeed Objects coders' accuracy on Emphasized dimensions was not significantly different from their accuracy on Unemphasized ($t = 0.67$).

Since the encoding process is so selective, it seems unnecessary to postulate that attention has a direct effect on perception.

Subsidiary Findings

Order of report.—The *S*s' preferred orders of report corresponded to their usual orders of encoding. When free to report in any order, Dimensions, Incomplete Dimensions, and Modified Objects coders reported the Emphasized dimension first over 70% of the time. Objects coders reported numbers first over 60% of the time.

In general, the dimension reported first was reported more accurately than those reported later. The largest differences between first and third reports occurred in the Incomplete Dimensions coders' reports of Unemphasized and of Equal dimensions. These reports were based, in part, on an unencoded visual trace, which fades more rapidly than an encoded item.

TABLE 3

ANALYSIS OF VARIANCE OF PERCENTAGE OF ERRORS WITH FORCED ORDER OF REPORT

Source	*df*	*MS*	*F*
Codes (C)	3	2675.73	3.63*
*S*s in C^a	29	737.95	
Instructions (I)	2	3036.76	13.85***
C × I	6	833.93	3.80**
*S*s in C × I^a	58	219.24	
Order of report (O)	2	1362.34	6.40***
C × O	6	195.05	
*S*s in C × O^a	58	212.80	
Left-right (L)	1	3465.75	17.77***
C × L	3	152.34	
*S*s in C × L^a	29	195.03	
I × O	4	61.38	
C × I × O	12	133.97	
*S*s in C × I × O^a	116	88.49	
I × L	2	186.63	3.27*
C × I × L	6	57.80	
*S*s in C × I × L^a	58	57.10	
O × L	2	147.59	
C × O × L	6	50.01	
*S*s in C × O × L^a	58	71.83	
I × O × L	4	81.93	
C × I × O × L	12	88.04	
*S*s in C × I × O × L^a	116	51.44	

Note.—Data obtained with One-Only instructions were not included in the analysis.

[a] Error term for preceding one or two terms.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Position of the stimulus.—Averaged across all codes and all instructions except One-Only, the rate of error on the left half of the stimulus was 10.9%; on the right, 16.0%. The greater accuracy on the left side was probably produced both by order of encoding and by order of report. All *S*s *encoded* the left value of each dimension before encoding the right value, according to their questionnaires. Furthermore, *E* observed that all *S*s checked off the left value first when they *reported* each dimension.

Speed of encoding.—Whereas 7 of the 16 *S*s who used a Dimensions coding strategy were usually unable to encode the entire stimulus (and are therefore labeled Incomplete Dimensions coders), none of the Objects coders reported such a difficulty. A plausible explanation is suggested by another unpublished study. When *S*s silently named pairs of cards, they took an average of 2.9 sec. per pair using Objects code, 4.2 sec. using Dimensions code (a difference significant beyond the .001 level). Since the visual image of a briefly presented stimulus fades considerably within 2 sec. (Sperling, 1960), the slower Dimensions coders are unable to encode it completely before it fades.

REFERENCES

BRUNER, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

CARMICHAEL, L., HOGAN, H. P., & WALTER, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.

CHAPMAN, D. W. Relative effects of determinate and indeterminate Aufgaben *Amer. J. Psychol.*, 1932, **44**, 163–174.

DEESE, J., & KAUFMAN, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180–187.

HERMAN, D. T., LAWLESS, R. H., & MARSHALL, R. W. Variables in the effect of language on the reproduction of visually perceived forms. *Percept. mot. Skills*, 1957, **7**, 171–186.

HUMPHREY, G. *Thinking: An introduction to its experimental psychology.* New York: Wiley, 1951.

KAY, H., & POULTON, E. C. Anticipation in memorizing. *Brit. J. Psychol.*, 1951, **42**, 34–41.

KUHLMAN, F. On the analysis of the memory consciousness: A study in the mental imagery and memory of meaningless visual forms. *Psychol. Rev.*, 1906, **13**, 316–348.

KÜLPE, O. Versuche über Abstraktion. *Ber. Kongr. exp. Psychol.*, 1904, **1**, 56–68.

LAWRENCE, D. H., & LABERGE, D. L. Relationship between recognition accuracy and order of reporting stimulus dimensions. *J. exp. Psychol.*, 1956, **51**, 12–18.

MILLER, G. A. The magical number seven, plus or minus two. *Psychol. Rev.*, 1956, **63**, 81–97.

MOORE, T. V. The process of abstraction. *U. Calif. Publ. Psychol.*, 1910, **1**, 73–197.

POLLACK, I. The assimilation of sequentially encoded information. *HumRRO Res. Lab. memo. Rep.*, 1952, No. 25.

RUBIN, E. Bericht über experimentelle Untersuchungen der Abstraktion. *Z. Psychol.*, 1913, **63**, 386–397.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74**(11, Whole No. 498).

SPERLING, G. A model for visual memory tasks. *Hum. Factors*, 1963, in press.

WILCOCKS, R. W. An examination of Külpe's experiments on abstraction. *Amer. J. Psychol.*, 1925, **36**, 324–341.

(Early publication received August 14, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 334–338

WAVE-LENGTH GENERALIZATION AS A FUNCTION OF SPACING OF TEST STIMULI[1]

HERBERT FRIEDMAN

Duke University

To find whether generalization curves depend upon the absolute or relative values of test stimuli, 15 pigeons were conditioned to 1 wave length of light and tested with 11 wave lengths spaced 5, 10, or 20 mμ apart. Each S was tested using each test series with intervening training. When responses were plotted against physical units, the gradients from the 3 test series were virtually identical. No support was found for a frame of reference process as predicted by the "units hypothesis."

In a recent review of stimulus generalization, Mednick and Freedman (1960) suggest that the extent of *S*'s response to a given test stimulus is proportional to the *number of stimulus units* that separate the training stimulus (with number of stimulus units defined in terms of the population of stimulus units *S* has experienced within the immediate experimental situation).

The "units hypothesis" suggests that stimulus generalization gradients are established in the course of testing as *S* learns the magnitude of the intervals separating test stimuli. It follows from the hypothesis that a generalization gradient obtained with a series of stimuli closely spaced in a narrow range would be steeper when plotted against physical units than one obtained using widely spaced stimuli over a larger range. However, these two gradients should be identical if graphed in terms of stimulus units. The view is suggestive of Lashley and Wade's (1946) position, and it also presents an analogy to Helson's (1959) adaptation level process, with the frame of reference in this case determined by the range of generalization-test stimuli and the intervals between them.

The units hypothesis by itself would readily explain the failure of Guttman and Kalish (1956) to find a relationship between generalization and discriminability for spectrum stimuli. It was expected that the slope of the generalization gradient would be inversely related to the size of the jnd, but this effect was not found. However, if *S*s responded on the basis of stimulus intervals, they would adjust to the particular degree of discriminability in the test region and the gradients produced by each group would therefore be similar in slope. Another problem of a methodological sort arises in experiments where additional test stimuli about the training stimulus are used to obtain a finer picture of the pattern of generalization (e.g., Guttman, 1956). If the units hypothesis is correct, *S* will adjust to the narrower stimulus spacing and no more detail will be seen. An inspection of such gradients, however, indicates that the form is much the same as would be expected if the added stimuli had not been used. Guttman (1956) also reports that typical wave-length gradients were obtained when six test stimuli were used on one side of the training stimulus and three on the other side, instead of using a balanced series. These two experiments suggest that pigeons are not strongly affected by stimulus spacing or range.

---

[1] This investigation was carried out during tenure as a Public Health Service Postdoctoral Research Fellow. The author wishes to thank Norman Guttman, who acted as sponsor and provided support from Grant M-3917, National Institute of Mental Health, United States Public Health Service.

The present experiment is specifically designed to determine if during color generalization testing pigeons respond on the basis of stimulus intervals, as suggested by the units hypothesis, or more directly to the wave length of the test stimulus. The *S*s are trained and then tested for generalization using stimuli 5 mμ, 10 mμ, or 20 mμ apart.

METHOD

Subjects.—The *S*s were 15 naive Carneau pigeons maintained at 75% ad lib. body weight by restricted feeding. A 23-hr. deprivation schedule was maintained during the experiment.

Apparatus.—A Skinner-type key pecking apparatus was used, the same as previously described by Guttman and Kalish (1956). During training of key pecking and testing, a Bausch and Lomb diffraction-grating monochromator illuminating the key was the sole source of light in the *S*'s chamber, except for the white light associated with operation of the food magazine. A white overhead light in the box provided illumination during magazine training.

Procedure.—Day 1 to Day 20 were consecutive.

Day 1: The *S*s were trained to eat from the food magazine, first with a light in the box and thereafter in the dark.

Day 2: The *S*s were conditioned to peck a key illuminated by monochromatic light 550 mμ in wave length. After training the pigeons were given 10 continuous reinforcements and then 15 reinforcements randomly during the next 75 responses.

Day 3: The pigeons were reinforced on a 30-sec. variable interval reinforcement schedule (VI-30 sec.) for 40 periods of 30-sec. duration separated by 10-sec. "blackouts" (no light supplied to the key).

Days 4–13: The same as Day 3 except that the mean interval between reinforcements was 1 min. (VI-1). Each scheduled reinforcement was kept available for a maximum of 10 sec. If the *S* did not make a response within this interval the scheduled reinforcement was canceled.

TABLE 1

ORDER OF PRESENTATION OF TEST CONDITIONS

Order	*N*	Spacing of Test Stimuli in Mμ		
		First Test	Second Test	Third Test
I	5	5	10	20
II	5	10	20	5
III	5	20	5	10

Day 14: Generalization testing was carried out. For all *S*s the test consisted of four random presentations under extinction conditions of each of 11 stimuli. For Cond. 5 mμ, the test stimuli ranged from 525 mμ to 575 mμ in 5-mμ steps. For Cond. 10 mμ, the test stimuli ranged from 500 mμ to 600 mμ in 10-mμ steps. For Cond. 20 mμ, the test stimuli ranged from 450 mμ to 650 mμ in 20-mμ steps. Five *S*s were randomly assigned to each condition.

Days 15–17: Training as on Day 4.

Day 18: Second generalization test; conditions assigned according to Table 1.

Days 19–21: Variable-interval training as on Day 4.

Day 22: Third generalization test, conditions assigned according to Table 1.

RESULTS

Test of units hypothesis.—The mean generalization gradients for the first test are shown in Fig. 1a with the mean percentage of the rate to 550 mμ plotted against the stimulus units and in Fig. 1b plotted against wave length. It is clear that the gradients are not similar when using stimulus units: the 5-mμ condition produces the flattest gradient, and the 20-mμ condition the steepest. Gradients in terms of wave length, however, are highly similar, with approximately equal responding to a given wave length without regard to the position of that stimulus within the three test series.

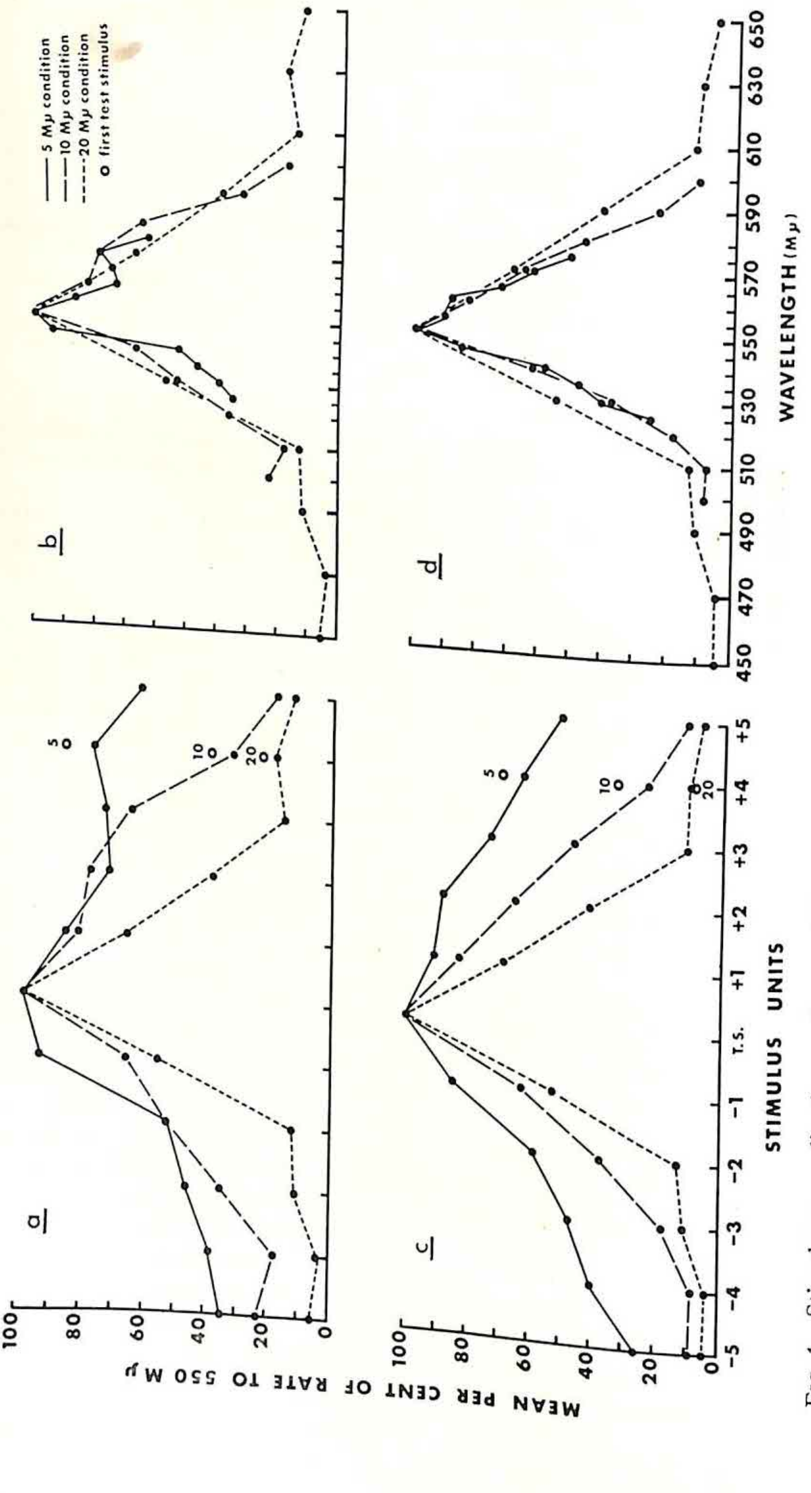


FIG. 1. Stimulus generalization gradients. (Figures 1a and 1b show the first generalization test, $N = 5$ in each group. Figures 1c and 1d are based on three tests with each *S* tested under each condition, $N = 15$. The mean responding to the first stimulus in each test is also shown.)

Figures 1b and 1c show the mean gradients for the total test series with each *S* tested under the three conditions. The picture is essentially the same as shown in Fig. 1a and 1b except for a greater regularity. In Fig. 1c, based on wave length, the 5-mμ and 10-mμ condition gradients are virtually identical while the 20-mμ curve is slightly flatter. Again, there are large and consistent differences among the gradients plotted against stimulus units. Each *S* gave its flattest gradient (in terms of units) under the 5-mμ condition, and 11 of the 15 *S*s gave their steepest gradients under the 20-mμ condition. A Treatments × Treatments × *S*s analysis of variance (Lindquist, 1953) of the results shown in Fig. 1c, based on stimulus units, indicated that the test conditions were significantly different ($p \leq .01$). The Conditions × Units interaction ($p \leq .01$) showed that the slopes of the gradients were significantly different.

Gradients based on absolute rate of responding.—Figure 2 shows the mean rate of responding given to the three wave lengths common to the three test conditions (530, 550, and 570 mμ) under the three test conditions based on $N = 15$. The slopes of the gradients are essentially identical, and agree with the gradients based on transformed data. A higher overall rate was found under the 20-mμ condition, presumably because the extreme stimuli in this series elicited little responding and consequently contributed little to the extinction of the response.

DISCUSSION

The present data unambiguously contradict the units hypothesis. It is clear that the *S*'s responding to a given stimulus is primarily determined by the distance in wave length between that stimulus and training stimulus, and that the effect of the test series is negligible.

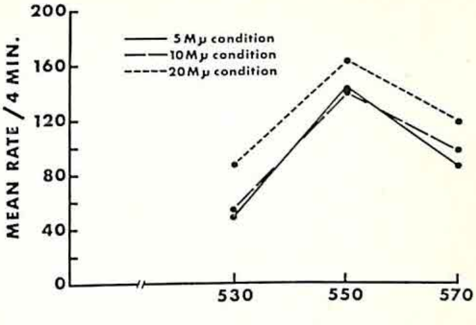


FIG. 2. Mean rate of responding based on three tests with each *S* tested under each condition, $N = 15$.

Since at the start of testing *S* has had no experience with the test series, the units hypothesis involves a specific prediction that *S* will respond to the first stimulus presented in the test series at approximately the rate given to the training stimulus. Figures 1a and 1c show that responding to the first test stimulus, which was always in the +4 units position, closely matches the average rate and is an orderly function of wave length. There is no evidence to suggest the development of a frame of reference during the course of testing.

The studies cited to support the units hypothesis employed human *S*s and frequency of sound (Hovland, 1937; Humphreys, 1939). Differences between these experiments and the present one may be due to differences either in the nature of the stimuli or of the *S*s. It may be that frame of reference hypotheses are inapplicable to qualitative continua like hue with which the stimuli are readily identifiable without resort to a frame of reference. The other possibility is that pigeons are not capable of using the information provided by the series of test stimuli (i.e., range and interstimulus spacing) to modify their responding. If this were the case, generalization gradients using pigeons would primarily reflect the training procedure, while gradients from human *S*s would be largely contaminated by the testing procedure. Some support for this last point can be obtained from Thomas and Hiss (1963) who used human *S*s and color stimuli in an experiment closely related to the present one. He found that the *S*s responded in part to both hue and stimulus spacing. Differences in method prevent a fuller comparison of the two experiments. A more detailed discussion of the importance of species differences in related problems is to be found in Kendler (1961).

References

GUTTMAN, N. The pigeon and the spectrum and other perplexities. *Psychol. Rep.*, 1956, **2**, 449–460.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HELSON, H. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill' 1959. Pp. 565–621.

HOVLAND, C. I. The generalization of conditioned responses: 1. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

HUMPHREYS, L. G. Generalization as a function of method of reinforcement. *J. exp. Psychol.*, 1939, **25**, 361–372.

KENDLER, T. S. Concept formation. *Annu. Rev. Psychol.*, 1961, **12**, 447–472.

LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

LINDQUIST, H. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MEDNICK, M. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

THOMAS, D. R., & HISS, R. A test of the "units hypothesis" employing wave-length generalization in human subjects. *J. exp. Psychol.*, 1963, **65**, 59–62.

(Received May 7, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 339–346

STIMULUS GENERALIZATION ALONG ONE AND TWO DIMENSIONS IN PIGEONS[1]

CHARLES M. BUTTER

University of Michigan

2 groups of pigeons were tested to compare 1-dimensional with 2-dimensional generalization and to determine whether stimulus dimensions have interacting effects of generalization. Following single stimulus training, both groups were tested for generalization to stimuli varied in wave length, in angular orientation, and in both dimensions. Both groups showed less generalization to stimuli varied in 2 dimensions than to stimuli varied in 1 dimension. Also, the 2 dimensions had interacting effects on generalization. An attempt was made to describe the mathematical form of this interaction in order to specify the function relating multidimensional to unidimensional generalization.

Although stimulus generalization has been investigated from many viewpoints, little attention has been directed toward the problem of how generalization takes place to stimuli varied along two or more dimensions as compared with generalization to stimuli varied in a single dimension. Experimental analysis of this problem may be considered one approach to the more general question of how different stimulus attributes combine to affect perceptual similarity. The experiments reported here were undertaken (*a*) to compare the amount of generalization to stimuli varied in one and in two dimensions and (*b*) to determine if two stimuli, each varied in a single dimension, exhibit independent or interacting effects on generalization when combined (Guttman, 1956). Evidence for independent effects of such stimuli on generalization would be provided if the generalization decrement to a stimulus varied in two dimensions were the sum of generalization decrements to stimuli varied in each dimension alone. On the other hand, stimulus interaction would result if generalization decrements to stimuli varied in one dimension did not summate. If such an interaction were found, it would be of further interest to attempt to determine its mathematical form and thus describe the function relating multidimensional to unidimensional generalization.

The method used in this study is the one employed by Guttman and Kalish (1956) to obtain wave-length generalization gradients in pigeons following training with a single spectral hue. The present study also employed wave length as one stimulus dimension and a visual-spatial dimension, angular orientation of a band of light (Butter & Guttman, 1957), as the second dimension.

Generalization to stimuli varied in one and in several dimensions has been investigated in two recent studies. Fink and Patton (1953) found that the amount of generalization decrement of an instrumental tube-drinking response in rats was directly related to the number of changed stimulus components. White (1958) found more generalization decrement in children to stimuli varied in two dimensions than to stimuli varied in either dimension alone. However, no conclusions were reached regarding the quantitative manner in which stimuli combined to affect generalization.

---

[1] The second experiment in this report forms part of a dissertation submitted to the Graduate School of Arts and Sciences, Duke University, 1959, in partial fulfillment of the requirements for the PhD degree in psychology. The author is indebted to Norman Guttman for his advice in planning these studies.

In the present study, two groups of pigeons were tested. The *S*s in Group 1 were tested for generalization to a small number of stimuli, while those in Group 2 were presented with a larger number of stimuli in order to obtain gradients of generalization.

METHOD

Apparatus.—The apparatus was an automatically-controlled Skinner box, designed for key pecking and similar to apparatuses used in other wave-length generalization studies (Guttman & Kalish, 1956). The special feature of the present apparatus, however, was that only a narrow band of light was transmitted to the key, and this band of light could be rotated about its center to provide various degrees of angular orientation. A circular translucent key, set within a frame, was masked so that a band running through its center, 1 in. in length and $\frac{1}{16}$ in. in width, was exposed to the light beam. The key could be rotated about its center within the frame by means of spur gears engaged by a rod. When *S*s were key pecking, the rod was moved back to disengage the gears on the key. The angular orientation of the band of light was set by means of a pointer attached to the other end of the rod. Settings, calibrated by a protractor, were accurate to 1°. Monochromatic interference filters (Bausch and Lomb Company), which could be mounted in a holder in the light path, had band widths varying between 7 and 9 mμ. The nominal and actual peak transmission values of the filters are shown in Table 1.[2] In subsequent reference to wave-length stimuli, nominal values will be used. However, in graphs, wave-length values are spaced according to their actual values. A wratten K-2 filter was used along with the 570- and 580-mμ filters in order to exclude lower wave lengths in the first-order spectrum.

TABLE 1

NOMINAL AND ACTUAL PEAK TRANSMISSION VALUES OF MONOCHROMATIC INTERFERENCE FILTERS

	Nominal Value (Mμ)	Actual Value (Mμ)
Group 1	520 550 580	522 548 579
Group 2	530 540 550 560 570	532[a] 539 548 561 571

[a] Upon completion of the experiment, spectrophotometric analysis indicated that the peak transmissions value of the 530-mμ filter had changed to 538 mμ. Because it is likely that this change occurred during the course of the experiment, responses to this wave-length stimulus are not included in the presentation of the data.

Subjects.—The *S*s were 37 experimentally naive white Carnean pigeons, 10 of which were assigned to Group 1, and 27 to Group 2. All *S*s were housed in individual cages and maintained at 75% of their ad libitum body weight throughout the experiment.

Procedure.—Following magazine training, all *S*s were trained to key peck by the method of successive approximations and were then given 50 continuous reinforcements on each of 2 successive days. During this stage of training, an overhead light was on in the apparatus, but in subsequent stages, the only illumination in the box was provided by the light from the key. In training sessions, the band of light on the key was illuminated by light of 550 mμ and was in a vertical position (90°). Following the completion of key peck training, *S*s were trained on a variable-interval (VI) reinforcement schedule with a mean interval of 1 min. Training sessions were administered over 10 days and consisted of 30 55-sec. light-on periods, during which the key was illuminated and key pecks were reinforced, alternating with 15-sec. light-off periods, during which no key pecks were reinforced.

On the day following the completion of VI training, *S*s were tested for stimulus generalization. Immediately before testing, *S*s were reinforced for pecking at the training stimulus for six 30-sec. periods. Subsequent generalization testing was carried out in extinction.

In generalization testing, *S*s in Group 1 were presented with stimuli differing in one and in two dimensions from the training stimulus by pairing three values of wave length, 520, 550, and 580 mμ, in all possible combinations with three values of angular orientation of the band of light, 40° (rotated 50° to the left of 90°), 90°, and 140° (rotated 50° to the right of 90°). In addition, 0° (horizontal) was paired with 550 mμ, so that a total of 10 different stimulus combinations was presented. These 10 stimulus combinations were presented 12 times, and within each of the 12 blocks, the stimulus combinations were arranged in a different random order. Each stimulus presentation lasted for 30 sec. and was followed by a 15-sec. dark period, during which stimuli were changed. On the following day, Group 1 *S*s received a second generalization test, administered in the same manner as on the first day of testing.

---

[2] Nominal values are those given in the manufacturer's catalogue, while actual values are those determined spectrophotometrically by the manufacturer and by *E*.

The *S*s in Group 2 were presented on the first day of generalization testing with all possible combinations of five values of wave length, 530, 540, 550, 560, and 570 mμ, and five values of angular orientation, 30°, 60°, 90°, 120°, and 150°. Each of these 25 stimulus combinations was presented six times, and within each of the six blocks, stimulus combinations were arranged in a different random order. Three blocks of stimuli were presented on the second day of generalization testing. In all other respects, testing was conducted in the same manner as with Group 1.

RESULTS

Group 1.—Mean responses to stimuli presented in the two generalization tests are plotted in Fig. 1 and 2 as a function of wave length, with angular orientation as the parameter, and as a function of angular orientation, with wave length as the parameter. It is evident that generalization on the two tests is highly similar, differing only in the overall response level. Responses on the two tests were pooled in analyses of the results.

Generalization decrement occurred to stimuli varied in either dimension; analysis of variance indicated that both wave length and angular orientation effects are highly significant (see Table 2). Further, there is an asymmetry in wave-length generalization; significantly more responses were given to 580 mμ than to 520 mμ, t (9) $= 7.90$, $p < .001$. Angular orientation generalization, on the other hand, was more nearly symmetrical; among the 10 *S*s, there was no consistent tendency to generalize more to one value of angular orientation than to the other.

In order to compare the amount of generalization to stimuli varied in one and in two dimensions, differences between mean responses to these two classes of stimuli were evaluated by a t test for correlated scores. The

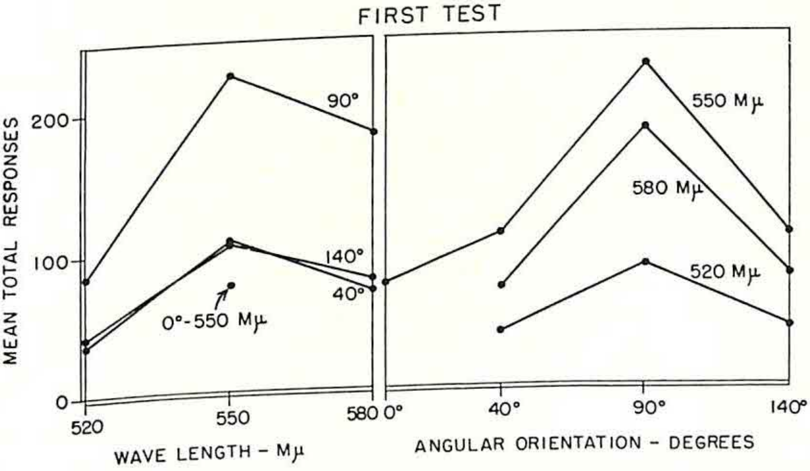


FIG. 1. Mean total responses to wave length and angular orientation stimuli on the first generalization test (Group 1).

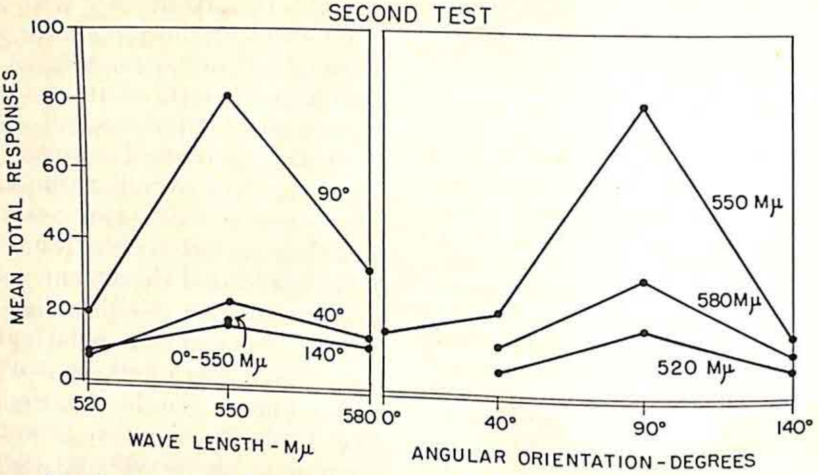


FIG. 2. Mean total responses to wave length and angular orientation stimuli on the second generalization test (Group 1).

results of this test showed significantly less generalization to stimuli varied in two dimensions than to stimuli varied in one dimension, $t\ (9) = 6.73,\ p < .001$.

The results may be further examined in order to determine whether stimuli on the wave-length and angular orientation dimensions exert independent or interacting effects on generalization. If the effects of these stimuli are independent, then the family of gradients along one dimension should be parallel. It appears, however, that the gradients are not parallel, but rather interact, as indicated by the significant Wave Length × Angular Orientation interaction (see Table 2). Furthermore, it appears from the graphs depicting the results that this interaction occurs in such a way that the gradients are multiples of each other, i.e., the form of the gradients suggests that relative generalization along one dimension is equal at different levels of the second dimension. Thus, for example, if Rt = responses to the training stimulus, Ra = responses to a stimulus varied only in angular orientation, Rw = responses to a stimulus varied only in wave length, and Raw = responses to a stimulus varied in both angular orientation and wave length, then:

$$\frac{\mathrm{Ra}}{\mathrm{Rt}} = \frac{\mathrm{Rwa}}{\mathrm{Rw}}$$

and

$$\mathrm{Rwa} = \frac{\mathrm{Ra} \times \mathrm{Rw}}{\mathrm{Rt}}$$

Dividing both sides of the equation by Rt:

$$\frac{\mathrm{Rwa}}{\mathrm{Rt}} = \frac{\mathrm{Ra} \times \mathrm{Rw}}{\mathrm{Rt} \times \mathrm{Rt}}$$

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL RESPONSES ON BOTH TESTS TO WAVE LENGTH AND ANGULAR ORIENTATION STIMULI: GROUP 1

Source	*df*	*MS*	*F*
Total	89		
Wave length (W)	2	109,751.5	37.75***
Angular orientation (A)	2	148,678.5	12.49**
W × A	4	13,058.6	5.79*
*S*s in groups	9		
W × *S*s	18	2,899.2	
A × *S*s	18	11,902.6	
W × A × *S*s	36	2,257.1	

Note.—Since there were repeated observations on *S*s, the significance levels of *F* ratios were tested by means of the conservative procedure suggested by Geisser and Greenhouse (1958).

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 3

MEAN RESPONSES OBTAINED AND PREDICTED BY THE MULTIPLICATIVE RULE: GROUP 1

Stimulus	Mean Obtained Responses	Mean Predicted Responses	PVE[a]
580 mμ, 40°	84.1	91.6	.92
580 mμ, 140°	89.5	86.5	.93
520 mμ, 40°	46.4	44.4	.90
520 mμ, 140°	49.3	41.9	.92

[a] Proportion of Variance Explained $= 1 - \frac{\sigma_{o.p}^2}{\sigma_o^2}$ where p = predicted responses and o = obtained responses (Levine, 1959).

According to the last equation, relative generalization to the stimulus varied in two dimensions equals the product of the values of relative generalization to the two stimuli each of which is varied in a single dimension.

Mean responses to stimuli varied in two dimensions and responses predicted by this "multiplicative rule" are shown in Table 3. It is evident that predicted responses closely approximate obtained responses; the overall mean error of prediction was 7.9%. Included in the same table are values of PVE (percent variance explained by predicted responses); the high values of this statistic indicate that responses of individual *S*s were predicted with a high degree of accuracy.

Group 2.—The results of generalization testing, shown in Fig. 3 and 4, are similar to those of Group 1. Analysis of variance indicates that both wave length and angular orientation effects were significant (see Table 4). Also, wave-length generalization was markedly asymmetrical, i.e., significantly more responses were given to 560 mμ than to 540 mμ, t (26) = 7.94, $p < .001$. As in the previous experiment no consistent

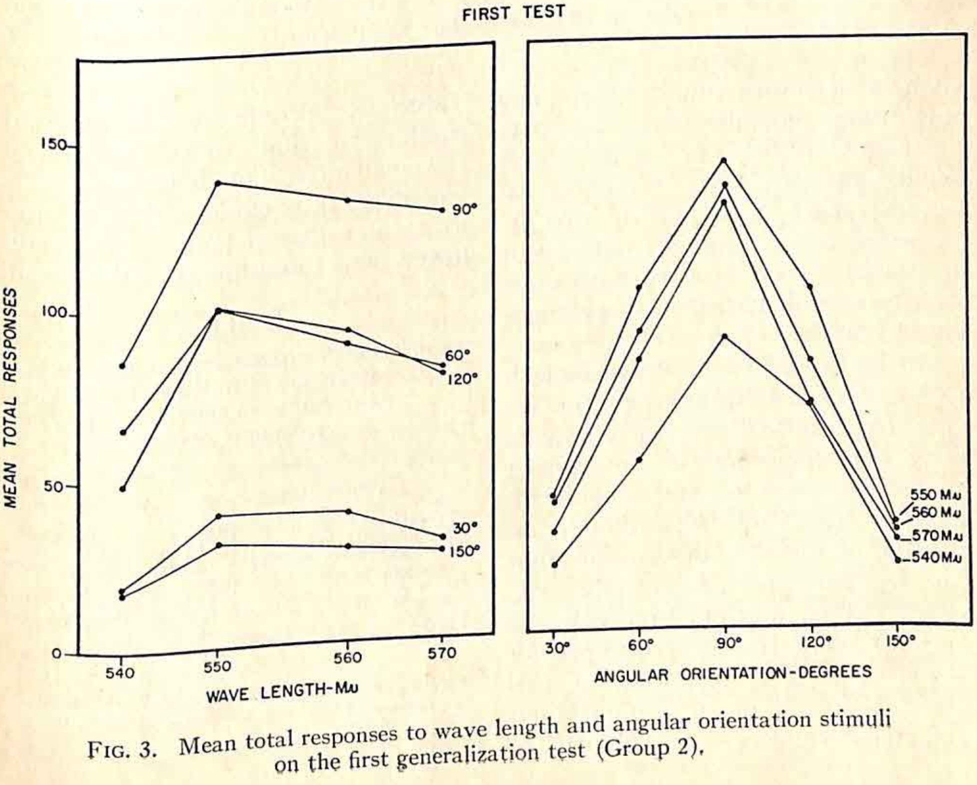


FIG. 3. Mean total responses to wave length and angular orientation stimuli on the first generalization test (Group 2).

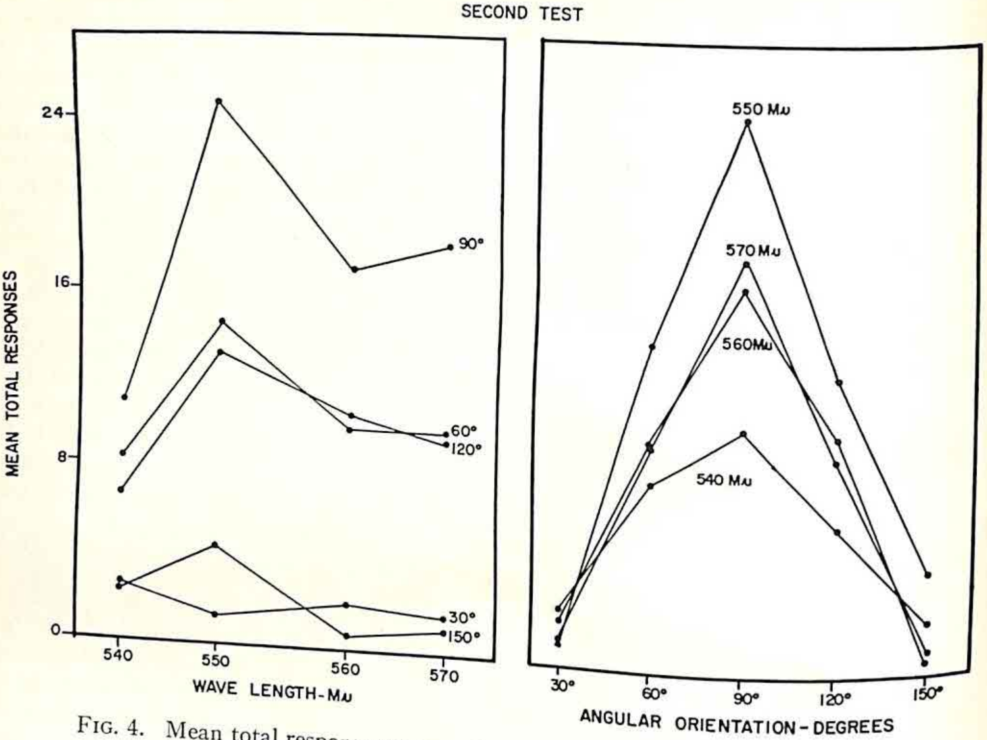


FIG. 4. Mean total responses to wave length and angular orientation stimuli on the second generalization test (Group 2).

asymmetries were found in angular orientation generalization. Further, significant differences were again found between mean responses to stimuli varied in one and in two dimensions; fewer responses were given to stimuli varied in two dimensions than to stimuli varied in one dimension, $t\ (26) = 7.54$, $p < .001$.

Finally, a significant Wave Length × Angular Orientation interaction results from the tendency for gradients along one dimension to broaden at successive levels of the second dimension (see Table 4), although inspection of Fig. 3 and 4 reveals that this tendency is not entirely consistent. The multiplicative rule described previously was used to predict responses to stimuli varied in two dimensions, and mean obtained and predicted responses are shown in Table 5. Prediction accuracy is similar to that found previously; the mean error of prediction was 9.7%. However, it is evident that the majority of values of PVE are markedly lower than those found in the results

TABLE 4

ANALYSIS OF VARIANCE OF TOTAL RESPONSES ON BOTH TESTS TO WAVE LENGTH AND ANGULAR ORIENTATION STIMULI: GROUP 2

Source	*df*	*MS*	*F*
Total	539		
Wave length (W)	3	35,477.4	9.35**
Angular orientation (A)	4	222,794.4	31.32***
Ss in groups	26		
W × A	12	3,403.4	3.27**
W × *Ss*	78	3,792.8	
A × *Ss*	104	7,114.3	
W × A × *Ss*	312	1,039.7	

Note.—The conservative procedure of Geisser and Greenhouse (1958) was used in evaluating the significance of the *F* ratios.
** $p < .01$.
*** $p < .001$.

TABLE 5

Mean Responses Obtained and Predicted by the Multiplicative Rule: Group 2

Stimulus	Mean Obtained Responses	Mean Predicted Responses	PVE
540 mμ, 30°	22.1	23.5	.46
540 mμ, 60°	57.6	68.1	.30
540 mμ, 120°	73.3	67.6	.39
540 mμ, 150°	21.5	21.7	.44
560 mμ, 30°	38.5	36.3	.56
560 mμ, 60°	87.2	104.9	.52
560 mμ, 120°	91.9	104.1	.74
560 mμ, 150°	28.9	33.4	.88
570 mμ, 30°	30.5	27.9	.90
570 mμ, 60°	89.3	80.7	.16
570 mμ, 120°	76.0	80.1	.52
570 mμ, 150°	27.2	25.7	.37

of Group 1, reflecting the poorer prediction of individual *S*s' responses.

Discussion

The results of this study indicate that the operant situation described by Guttman and Kalish (1956) in their study of stimulus generalization can also be employed to obtain gradients along two dimensions from individual *S*s. Moreover, these gradients are similar to those generated in tests in which stimuli are varied only in a single dimension. Thus, in both, the wave-length gradients around 550 mμ were asymmetrical in the same direction as those reported in previous studies (Blough, 1961; Guttman & Kalish, 1956; Hanson, 1959), while angular orientation gradients were more symmetrical, resembling those described earlier for this dimension (Butter & Guttman, 1957).

The results of both groups demonstrate that generalization to stimuli varied in two dimensions is less than generalization to stimuli varied in either dimension alone. This finding is consistent with previous ones (Fink & Patton, 1953; White, 1958) and suggests that perceptual similarity can be altered more readily by multidimensional than by unidimensional stimulus changes.

A significant interaction between dimensions was found in the results of both groups. This finding is not consistent with the view that multidimensional generalization is the result of the algebraic summation of generalization decrements to all discriminable stimulus changes. The form of the generalization gradients obtained in Group 1 suggested, rather, that relative generalization to a stimulus varied in two dimensions equals the product of the values of relative generalization to stimuli varied in a single dimension alone. In perceptual terms, this multiplicative rule states that the similarity between stimuli differing in several dimensions is a joint function of the perceived similarity between stimuli within each of the relevant dimensions. The multiplicative rule predicted mean responses and responses of individual *S*s in Group 1 with a high degree of success. However, predictions of individual *S*s' responses in Group 2 were much less accurate, a finding which may be accounted for by the fact that each stimulus was presented less frequently to these *S*s than to those in Group 2, thereby tending to reduce reliability of measurement. It is also possible that response interactions between stimulus presentations contributed variability to the results and thus to predictions of responses. The lower accuracy of predictions of the Group 2 data may also be due, at least in part, to lack of complete orthogonality of the dimensions, i.e., probabilities of response to each aspect of the stimulus may not be independent. However, since no constant errors of prediction were found in mean responses or in responses of individual *S*s, it appears unlikely that departure from orthogonality had any significant effect on the results. Finally, one should recognize the possibility that multidimensional generalization cannot be described by any simple quantitative rule that is applicable to a variety of stimulus dimensions and experimental conditions. Future studies using other dimensions and methods of testing may provide more information about the manner in which multidimensional generalization occurs.

REFERENCES

BLOUGH, D. S. The shape of some wavelength generalization gradients. *J. exp. Anal. Behav.*, 1961, **4**, 31–40.

BUTTER, C. M., & GUTTMAN, N. Stimulus generalization and discrimination along the dimension of angular orientation. *Amer. Psychologist*, 1957, **12**, 449. (Abstract)

FINK, J. B., & PATTON, R. M. Decrement of a learned drinking response accompanying changes in several stimulus characteristics. *J. comp. physiol. Psychol.*, 1953, **46**, 23–27.

GEISSER, S., & GREENHOUSE, S. W. An extension of Box's results on the use of the *F* distribution in multivariate analysis. *Ann. math. Statist.*, 1958, **29**, 885–891.

GUTTMAN, N. The pigeon and the spectrum and other perplexities. *Psychol. Rep.*, 1956, **2**, 449–460.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HANSON, H. M. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58**, 321–334.

LEVINE, M. A model of hypothesis behavior in discrimination learning set. *Psychol. Rev.*, 1959, **66**, 353–366.

WHITE, S. W. Generalization of an instrumental response with variations in two attributes of the CS. *J. exp. Psychol.*, 1958, **56**, 339–343.

(Received May 14, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 347–351

ASSOCIATION VALUE, FAMILIARITY, AND PRONUNCIABILITY RATINGS AS PREDICTORS OF SERIAL VERBAL LEARNING

RICHARD H. LINDLEY[1]

Trinity University

3 measures of meaningfulness were compared as predictors of the ease of serial verbal learning of 12-item lists of consonant-vowel-consonant trigrams of 30–80% association value (AV). There were 4 main groups of Ss; the lists had the same values for 2 measures of meaningfulness and 1 which was different (e.g., equal average AV and familiarities but different average rated pronunciabilities). Ss were 80 men and women college students. The results were that ease of learning was not a function of rated pronunciability or familiarity ($p > .05$) but was a function of AV for the number of correct responses measure of learning ($p = .05$).

Three related measures of the meaningfulness of trigrams are association value (AV), familiarity, and pronunciability rating (PR). There has been considerable recent interest in determining which of these measures most accurately predicts the ease of verbal learning and hence presumably underlies the dimension of meaningfulness (Archer, 1961; Di Mascio, 1959; Johnson, 1962; Lindley, 1960; Saltz & Ager, 1962; Terwilliger, 1962; Underwood & Schulz, 1960). The present study compared these three measures as predictors of serial verbal learning with moderately meaningful CVC (consonant-vowel-consonant) trigrams.

METHOD

Experimental design.—There were four main groups of *S*s. Group 1 learned a list of low AV, high PR (i.e., hard to pronounce) unfamiliar trigrams; Group 2 learned a list of low AV, low PR (easy to pronounce), unfamiliar trigrams; Group 3 learned a list of high AV, low PR, familiar trigrams; Group 4 learned a list of high AV, low PR, unfamiliar trigrams. A significant difference in ease of learning between Groups 1 and 2 may be attributed to the difference in PRs; between Groups 3 and 4 to the difference in familiarity; and between Groups 2 and 4 to the difference in AVs.

Materials.—Table 1 presents the lists of trigrams that were used. In addition to the two different lists of items for each main group, there were two different serial orders of each of the eight lists. A few items were used in both lists of a given group; the extent of this duplication can be determined from Table 1.

One estimate of pronunciability was the Underwood and Schulz (1960, Appendix E) PRs. In these ratings *S* was instructed to rate the items on a nine-point scale according to their relative ease or difficulty of pronouncing. Low numbers indicate easy to pronounce items; high numbers indicate hard to pronounce items.

A second estimate of pronunciability was obtained from 30 *S*s from Trinity University who rated the pronunciability of all items used in this study. These ratings were obtained after the serial learning data had been collected; none of these *S*s participated in the serial learning study.

Both the recent Archer (1960) and Noble (1961) AVs were used to estimate AV. Noble's (1961) association ratings (a' values) were also used; in these ratings *S*s used a five-point scale to rate the number of associations they thought the items would elicit.

---

[1] The author would like to express his appreciation to Roy E. Hardman for his very capable assistance in all phases of this research.

TABLE 1

LISTS USED

Group 1		Group 2		Group 3		Group 4	
ZOJ	CYR	VIF	VAD	CED	ZAM	KUS	BEK
CYR	GUD	ZUN	MEF	HAZ	HOV	NEL	JAS
XAT	XAN	YAG	YAG	VER	VIL	JAS	PUK
COH	YIR	TOZ	PEX	FAC	JUS	KER	RYM
NIQ	QAD	MEF	TIV	BON	WOM	CUZ	KUF
JUX	COH	CIB	ZON	DIR	REC	LYN	WEY
QAD	BUV	HEG	WIB	CAK	WIF	PAX	BIM
YUW	RAJ	KIV	TUD	BOT	JOK	BOZ	GOP
RAJ	YUQ	PEX	JOR	JAD	DAL	KAV	YUM
GUD	GIH	VAD	VIF	ROZ	SUK	DUX	BAH
WIH	TUW	YIN	MOG	VAL[a]	ROX	YAH	FEY
ROQ	ZOX	ZED	JEC	DOK	MAK	CIL	WUN

[a] This item was used by mistake; Underwood and Schulz PRs are not available for it.

In the Noble *a′* values low numbers indicate few rated associations; high numbers, many.

Frequency of occurrence was estimated by using the Underwood and Schulz (1960, Appendix D) Total (frequency) Count. This count gives the estimated frequency of occurrence of trigrams per 141,000 words. Familiarity was estimated by determining whether or not an item was the first three letters of one or more words which appeared in English text 10 times per million or more as reported by Thorndike and Lorge (1944).

The various indicators of meaningfulness as they apply to the groups used in this experiment are presented in Table 2. The items were selected for each list with the restriction that formal intralist similarity (number of repeated letters) be roughly equal in all lists.

Procedure.—The items were presented on a memory drum at a 2-sec. rate with a 14-sec. rest between trials. An asterisk served as a cue for the first item. Typical instructions for learning by the anticipation method when the items are spelled were read to *S* at the start of the experiment.

The lists were learned to a criterion of one perfect recitation or for 25 trials, whichever took longer. Thus data for the number of trials to reach the criterion and for the number of correct responses during 25 trials were collected. One exception was made to the above criterion; no *S* was tested for more than 60 trials. There was 1 *S* in Group 1, 4 in Group 2, and 1 in Group 3 who were tested for 60 trials.

Subjects.—The *S*s were 80 men and women students from Trinity University. None of the *S*s had served previously in a memory drum learning experiment. There were 16 different lists; *S*s were randomly assigned to experimental conditions in blocks of 16 such that each *S* within a block learned a different list.

TABLE 2

MEAN MEANINGFULNESS VALUES FOR THE GROUPS

	Groups			
	1	2	3	4
Underwood and Schulz PRs	5.35	3.36	3.16	not available
Trinity PRs	5.80	3.70	3.23	3.16
Archer AVs	33.83	34.88	76.96	78.63
Noble AVs	48.32	55.38	88.83	83.59
Noble *a′* values	1.72	1.86	3.02	2.57
Underwood and Schulz frequencies	0.75	25.42	56.79	8.63
Average number of familiar items (Thorndike-Lorge count)	0.00	0.50	9.00	0.50

Note.—See text for explanation.

TABLE 3

Means and *SD*s for Trials to Reach Criterion, for Number of Correct Responses in 25 Trials, and for Errors per Trial

Groups	Trials		Number Correct		Intrusion Errors per Trial		Invention Errors per Trial	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	37.65	12.33	114.55	37.86	.50	.32	.67	.44
2	38.20	13.34	112.70	41.84	.62	.28	.78	.47
3	26.95	11.67	156.25	38.96	.75	.34	.39	.32
4	33.40	12.77	141.95	44.40	.73	.42	.44	.22

Results and Discussion

Serial learning.—Table 3 presents the means and *SD*s for the number of trials to reach criterion and for the number of correct responses during 25 trials. The analyses of variance for these measures are presented in Table 4. In both analyses meaningfulness was the only significant source of variance; the Duncan multiple range test (with $\alpha = .05$) as described by Edwards (1960) was used to test the differences between the main groups. These analyses are presented in Table 5. Of the three measures of meaningfulness used, only AV accurately predicted ease of learning and then only when a number of correct responses measure of learning was used. However, 20% of the *S*s in Group 2 did not reach criterion within the limit of 60 trials as compared to the *S*s in Group 4, all of whom reached criterion within 60 trials. Since *S*s who did not reach criterion were given a score of 61 (the earliest trial on which they could have reached criterion), the mean number of trials to criterion for Group 2 likely are biased by being too low; this bias could account for the failure to find a significant difference in trials between the groups that differed in AV (Groups 2 and 4).

TABLE 4

Analyses of Variance of the Serial Learning Data

Source	Trials to Criterion			Number Correct in 25 Trials	
	df	*MS*	*F*	*MS*	*F*
Meaningfulness	3	540.10	3.26*	9082.88	5.17**
Lists	4	110.68	.62	1851.79	1.02
Orders	8	101.00	.56	1385.64	.76
Replications	4	144.98	.81	1464.80	.81
Remainder	60	179.18		1818.11	
Pooled error term[a]	76	165.55		1755.76	

[a] This is the appropriate error term for testing the effects of meaningfulness.

* $p < .05$.

** $p < .01$.

The serial learning data were analyzed into two stages as a test of the two-factor theory of verbal learning of Underwood, Runquist, and Schulz (1959). According to these authors the first stage consists of response learning akin to the response integration notion of Mandler (1954) and the second stage of associative learning in which associations are formed between the integrated items. It is assumed that the two stages can proceed concurrently. Postman (1961) wrote:

> The duration of the integrative stage is estimated by the number of the trial on which the correct response is first given anywhere in the list. The length of the associative stage is measured by the number of trials between the end of the integrative stage and the first occurrence of the correct sequential association (p. 101).

This analysis is presented in Table 6.

Analysis of variance shows that

TABLE 5

Differences between Group Means in Serial Learning

Groups	Trials to Criterion				Groups	Number Correct in 25 Trials			
	Groups			Shortest Significant Ranges		Groups			Shortest Significant Ranges
	1	4	3			1	4	3	
2	0.55	4.80	11.25	8.83	2	1.85	29.25	43.55	28.73
1		4.25	10.70	8.54	1		27.40	41.70	27.80
4			6.45	8.12	4			14.30	26.42

there were no significant differences as a function of replications, orders, or lists in the mean number of trials on which the correct responses were first given ($F = .49, .63, 1.43, df = 4/60, 8/60, 4/60$, respectively). Meaningfulness was a significant source of variance, $F\ (3, 76) = 6.62,\ p < .01$. The Duncan multiple range test ($\alpha = .05$) showed that the only significant difference was between Groups 3 and 4 vs. Groups 1 and 2. As predicted by the two-stage theory, the high AV items were more available to *S*s and were given in significantly fewer trials than were the low meaningfulness items. No comparable differences occurred for the associative stage. The mean number of trials separating the first appearance of the response from the first correct placement was for Group 1, 1.40; for Group 2, 1.47; for Group 3, 1.21; and for Group 4, 1.37. Analysis of variance shows no significant differences among these means, $F\ (3, 60) = .40$.

TABLE 6

Means and *SD*s for the Trial Number on Which Responses Were First Given and First Placed Correctly

Groups	First Given		First Placed	
	Mean	*SD*	Mean	*SD*
1	11.77	5.15	13.17	4.93
2	11.91	4.59	13.38	5.12
3	6.87	2.31	8.08	2.49
4	8.61	4.14	9.98	4.17

Error data.—The means and *SD*s for the number of three-letter errors per trial made during the trials to reach criterion are presented in Table 3. An intrusion is a misplaced response from the list; an invention is a response which does not occur in the list.

Analysis of variance showed that there were no significant differences in intrusion errors as a function of replications, orders, lists, or meaningfulness ($F = 2.24,\ .39,\ .50,\ 2.18,\ df = 4/60, 8/60, 4/60, 3/76$, respectively, $p > .05$). With respect to the invention type of error, replications and orders were not significant ($F = .51,. 51, df = 4/60, 8/60$, respectively, $p > .05$); lists and meaningfulness were significant ($F = 3.02, 4.92,\ df = 4/60,\ 3/60$, respectively, $p < .05$). Since lists were significant for invention errors, the significant effect of meaningfulness must be severely qualified. The most justified conclusion appears to be that meaningfulness did not have a clear-cut effect on either type of three-letter errors.

The serial learning results of the present study do not confirm the findings of Underwood and Schulz (1960) and Underwood and Postman (1960) who found that ease of paired-associate and serial learning was closely related to

PRs. The discrepancy in these results does not appear to be related to any unreliability of the PRs of the trigrams used in this study since the mean ratings of the lists used here (determined by 30 *S*s in the present study) are quite similar to those reported by Underwood and Schulz (1960). Of course, the range of meaningfulness in this study was narrower than that of the other studies and the present study used only CVC trigrams while the other studies had more than one type of trigram. Either or both of these differences in procedure could possibly account for the discrepancy in results.

For both measures of serial learning in this study, the familiar lists were learned better than the unfamiliar ones; however, familiarity by itself is not a significant source of variance. Thus the results of the present study do not confirm the previous assertion of Lindley (1960) that familiarity accurately predicts ease of serial verbal learning independently of AV. The previous Lindley (1960) study used the out-of-date Glaze (1928) and Krueger (1934) AVs—the recent Archer (1960) and Noble (1961) AVs had not appeared when Lindley's (1960) study was published; the unreliability of the Glaze and Krueger AVs may well account for the Lindley (1960) results (cf., Archer, 1961).

The stage analysis strongly supports the Underwood et al. (1959) two-stage theory of verbal learning. The stage analysis suggests that the effect of meaningfulness was on the response integration phase of verbal learning and that meaningfulness had no effect on the associative stage. These findings are quite comparable to results obtained by Postman (1961) with the learning of serial lists of meaningful words.

REFERENCES

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

ARCHER, E. J. Supplementary report: Familiarity or low validity? *J. exp. Psychol.*, 1961, **62**, 207–208.

DI MASCIO, A. Learning characteristics of nonsense syllables: A function of letter frequency. *Psychol. Rep.*, 1959, **5**, 585–591.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

JOHNSON, R. C. Reanalysis of "Meaningfulness and Verbal Learning." *Psychol. Rev.*, 1962, **69**, 233–238.

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.

LINDLEY, R. H. Association value and familiarity in serial verbal learning. *J. exp. Psychol.*, 1960, **59**, 366–370.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

NOBLE, C. E. Measurements of association value (a), rated associations (a') and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychol. Rep.*, 1961, **8**, 487–521.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61**, 97–110.

SALTZ, E., & AGER, J. W. Issues in scaling meaningfulness: Noble's revised CVC norms. *Psychol. Rep.*, 1962, **10**, 25–26.

TERWILLIGER, R. F. Note on familiarity and verbal learning. *Psychol. Rep.*, 1962, **10**, 409–410.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intra-list similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received May 14, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 352–361

COMPARISON OF SERIAL AND PAIRED-ASSOCIATE LEARNING[1]

LEONARD M. HOROWITZ AND CHIZUKO IZAWA

Stanford University

Serial learning has been viewed as a paired-associate (PA) task where a single item is the effective stimulus for the following item. Recent evidence suggests that this view fails to describe the serial task adequately. To help understand this evidence, 4 lists were composed which varied in their internal association. The interitem associative strength was constant in 3 lists, and the strength between adjacent items was varied. In a 4th list no 2 items were associates. 4 groups of 20 Ss learned a list both serially and by the PA method. The PA items were formed from adjacent items of the serial list. Group S-P learned the serial list 1st; Group P-S, the PA list 1st. Comparisons were made between the 2 tasks and between the 2 directions of transfer. When adjacent items were strongly associated, the amounts of transfer between tasks were about equal. When remote items were associated, greater transfer occurred from PA to serial learning, replicating earlier findings. When no items were associated, greater transfer occurred from serial to PA learning.

According to an assumption that is often made in studies of serial learning, stimulus-response associations develop between adjacent items of the serial list. The serial task A-B-C-D has thus been viewed as a paired-associate (PA) task which contains the S-R pairs A-B, B-C, and C-D. This "chaining hypothesis" ignores other features of the serial list, and recent evidence has suggested that it fails to describe the effective stimulus adequately. The present study is designed as one step towards clarifying the nature of the effective stimulus. Serial and PA learning are compared on lists which vary in their internal association, and the amounts of transfer between tasks are also compared.

Primoff (1938) presented *S* with a list of items to be learned serially. Then *S* learned a PA list whose S-R pairs were formed from adjacent items of the serial list. Every Item *n* of the serial list became paired with Item $n + 1$, the response. Primoff concluded that this procedure impaired *S*'s PA performance, but the comparison was not methodologically sound (Umemoto & Hilgard, 1961; Young, 1961). Young (1959) refined the design of Primoff's study as follows. Group S-P learned a serial list first and then a PA list derived from it by Primoff's method; Group P-S learned the lists in reverse order. To assess the amount of transfer from one task to the other, the performances of these groups were compared in serial learning and in PA learning. The transfer from PA to serial learning was considerable, while that from serial to PA learning was negligible.

Thus, when *S* learns associations by the PA method, he seems able to use these associations to reconstruct the serial list. However, if he learns the serial task first, he has little advantage in then learning the PA task. This result implies that the chaining hypothesis is not valid. And, at least for the lists used by Primoff and Young, an analysis of the serial task into simple S-R pairs is not warranted.

---

[1] This research was supported in part by Grant M-5428 of the United States Public Health Service. The article is based upon a thesis submitted by the junior author to Stanford University in partial fulfillment of the requirements for the degree of Master of Arts.

Recent evidence by Horowitz (1961) suggests that the effective stimulus in serial learning may depend upon the interitem relationship within a list. In his study *S*'s performance in serial learning was measured by the reconstruction method. After *S* observed the items in a standard order, he received a packet of small cards containing one item each. The *S* reordered these cards to match the standard order of presentation, and his performance was scored in two ways. The "correct placement score" measured the number of items that *S* placed in correct serial position, and the *P* score measured the number of pairs of items that were placed correctly relative to one another. If *S* used serial position as a stimulus in serial learning (and his *P* score improved thereby), an expected relationship could be predicted between the two measures. This relationship was found when intralist similarity was high. But when intralist similarity was low, the *P* score was above that expected from the correct placement score, and *S* must have formed associations between adjacent items more often.

These results suggest that the chaining hypothesis may be valid during serial learning if the serial items are distinctive. Otherwise, two sources of interference may prevent *S* from using such a strategy. For one thing, to form these associations when the items are highly similar implies that similar stimuli must elicit similar responses, and considerable response competition can be expected (Osgood, 1949). By the same principle, backward associations would also compete with the forward associations (Murdock, 1958). Thus, to minimize these sources of interference, *S* would have to restructure the serial task, say, by learning each item as a response to its serial position or to a cluster of preceding items. In the Primoff and Young studies, the items to be learned were all CVC trigrams, three-letter words, or two-syllable adjectives. The interference among such items may have been too great for *S* to learn them by a simple chaining procedure; and for this reason, positive transfer did not occur from serial to PA learning.

Thus, the chaining hypothesis may still be valid if the items are sufficiently distinctive. If so, several consequences should follow. First, there should be positive transfer from the serial task to the PA task derived from it; likewise from the PA task to the serial task. Second, the tasks should be learned at similar rates. And third, similar distributions of errors should occur for both tasks. If these implications are disconfirmed, the effective stimulus of the serial task must embody additional features of the serial list.

The present study was designed to test these implications when the interference among items within a serial list varies. Four lists of words were prepared which varied in the relationship among items. The interitem relationship was defined by the associative strength between words.

METHOD

Materials.—Twenty-one words were selected from the Minnesota norms (Russell & Jenkins, 1954) of the Kent-Rosanoff Word Association Test. Each word occurred more than 10 times per million in the Thorndike-Lorge (1944) frequency count. The interitem associative strength, as defined by Deese (1959, 1961), was 38.4%. The words were so selected that each word elicited some one word of the list more strongly than it did any of the other words. These words were: BUTTER, CHEESE, YELLOW, RED, BLACK, WHITE, BEAUTIFUL, WOMAN, CHILD, BOY, ROUGH, STREET, LAMP, TABLE, CHAIR, HARD, SOFT, LIGHT, HOUSE, DARK, and WINDOW.

The words were arranged in three different orders to make Lists F (forward), B (backward), and R (remote). The above order was adopted as the standard order of presentation for List F. In this arrangement the mean forward associative strength between adjacent items was 15.1%, the maximum value possible for these words. The mean backward associative strength between adjacent items was 8.1%. The remaining associative strength of the list (38.4%—15.1%—8.1%) approximately equals the summed associative strengths between a given item and nonadjacent items, averaged across the words.

The order of the words in List B was the reverse of that listed above. This particular arrangement was selected to test an assumption that has appeared in the recent literature explaining why a serial list is easier to learn than the PA list derived from it. It has been assumed (e.g., Young, 1959) that interfering backward associations become established during PA learning but not during serial learning. In List B the mean backward associative strength between adjacent items was 15.1% so strong backward associations must be present during both serial and PA learning. Hence, if this assumption is valid, *S*'s performance in serial learning should more nearly equal that in PA learning for List B.

The same words were arranged in a third order as follows to form List R: RED, BOY, STREET, BLACK, SOFT, LAMP, ROUGH, WOMAN, CHAIR, CHILD, TABLE, WHITE, HARD, CHEESE, LIGHT, BUTTER, BEAUTIFUL, HOUSE, YELLOW, WINDOW, and DARK. In this arrangement the mean strength of the forward and backward associations between adjacent items is 0, and all of the associative strength of the list is between nonadjacent items.

A fourth list, List I (independent), contained words that were not associates of one another (Deese, 1959). These words appeared in the following order: NEEDLE, TONIGHT, MOUNTAIN, AFRAID, BREAD, THIEF, FORGET, COLD, JUMP, WOOD, SCHOOL, MOUTH, CARPET, STOP, FAMILY, SALT, BIG, TURTLE, BLOSSOM, RIDE, and LAKE. They all occurred more than 10 times per million in the Thorndike-Lorge frequency count, and their interitem associative strength is 0.

Experimental procedure.—Each *S* was presented with a practice list first. The practice list contained eight two-syllable adjectives to be learned by the serial anticipation method to a criterion of one perfect trial. These items were: CUTTING, EAGER, POLITE, AWKWARD, MORBID, INTENSE, LEGAL, and WEARY.

After a brief rest, half of the *S*s were placed in Group S-P and the remaining *S*s in Group P-S. The *S*s of Group S-P participated in a serial task and 24 hr. later in a PA task. Those of Group P-S participated in the PA task first and 24 hr. later in the serial task. Each group was further subdivided into four subgroups, and a different list was assigned to each. Therefore, there were eight experimental groups, and within each group *S* had two tasks. Primoff's method was always used to derive *S*'s PA list from his serial list. Thus, there were 21 items in the serial task and 20 pairs of items in the PA task.

The serial list was presented on a Lafayette memory drum at a 2-sec. rate with 2 sec. between items; the intertrial interval lasted 6 sec. The *S* was instructed to anticipate each word before it appeared on the drum. This procedure was repeated until *S* could anticipate all the items correctly.

Each stimulus-response pair in the PA task was presented on a 5 × 8 in. index card. These cards appeared behind a 4 × 12 in. cardboard screen with an aperture $\frac{3}{4} \times 3$ in. The aperture permitted *E* to expose the stimulus alone for 2 sec. and then, by moving the index card, to expose the response alone for 2 sec. There was no delay between stimulus and response presentations, but the delay between pairs was approximately 2 sec. The intertrial interval lasted approximately 10 sec. After all the items had been presented once, *S* was instructed to anticipate the response during the 2-sec. stimulus presentation. The pairs were presented in randomized order on each trial, and the procedure was repeated to a criterion of one perfect trial.

Subjects.—Eighty *S*s, 10 in each condition, participated in the experiment. (Although 85 *S*s were tested, 2 *S*s failed to reach criterion in PA learning, and 3 others failed to return for the second session. Of those who did not reach criterion, 1 had been assigned to List R and the other to List I.) These *S*s were all students enrolled in the introductory psychology class at Stanford University, and none of them had had previous experience in verbal learning experiments.

RESULTS

Practice task.—The mean number of trials needed to reach criterion

TABLE 1

Mean Number of Trials to Reach Criterion in Each Condition

List	Serial Learning		PA Learning	
	Group S-P	Group P-S	Group P-S	Group S-P
F	4.2	2.1	9.6	7.2
B	7.8	3.6	12.6	10.5
R	8.7	2.8	16.6	12.3
I	8.9	5.2	18.4	7.5

on the practice task was computed for each of the eight groups. These means ranged from 3.8 to 5.5, and an analysis of variance showed that they did not differ significantly ($F < 1.00$). Thus, the eight groups of *S*s can be considered initially comparable.

Acquisition.—Table 1 presents the mean number of trials needed to reach criterion on each learning task by *S*s of each condition. An analysis of variance revealed significant differences between the two tasks, F (1, 72) = 129.04, $p < .001$ and between the four lists, F (3, 72) = 5.89, $p < .01$. Two different error terms were used in this analysis since the experimental design contained comparisons between *S*s and within *S*s (Lindquist, 1953, pp. 281–284). For both error terms, however, $df = 72$. The difference between the two tasks was no greater for List F than for List B, and backward associations seem to play a minor role in producing the difference between tasks.

On the average, the lists can be rank-ordered as follows in increasing difficulty: F, B, R, and I. This order of difficulty held true of the first task, whether it was serial or PA learning. On the second task, though, the order of difficulty changed since the amount of transfer was greater for some conditions than for others. For this reason, *S* could learn List I by the PA method more efficiently if he learned the serial task first and then transferred to the PA task (a total of $8.9 + 7.5 = 16.4$ trials); to learn the PA task directly required 18.4 trials.

Thus, the order in which the tasks were presented interacted significantly with the nature of the task, F (1, 72) = 61.90, $p < .001$; also, the nature of this interaction depended upon the particular list used, thereby producing a significant List × Task × Order interaction, F (3, 72) = 3.93, $p < .05$. No other main effect variable or interaction was significant.

Order as a variable, of course, is necessarily confounded with other variables. For example, the amount of forgetting that occurs between tasks may depend upon the task learned first. Also, *S* may "learn how to learn" more efficiently with one task than with the other. Hence, it is only in the most conservative, operational sense that the present authors discuss differences between orders and interactions with order. Clearly, though, the advantage that occurs in serial or PA learning as a second task (for whatever reason) depends upon the interitem relationship within the list. This point is elaborated below, where transfer scores are examined in detail.

Transfer scores.—The serial scores achieved by Group P-S were converted to serial transfer scores, $T_{PA}(S)$, and the PA scores achieved by Group S-P were converted to PA transfer scores, $T_S(PA)$. A transfer score is defined as a measure of the amount, in percentage units, that an experimental *S*'s score surpassed that of the average control *S* performing the same task on the same trial. Serial transfer scores were computed by the following formula:

$$T_{PA}(S) = \frac{S_{P-S} - \bar{S}_{S-P}}{\bar{S}_{S-P}} \times 100.$$

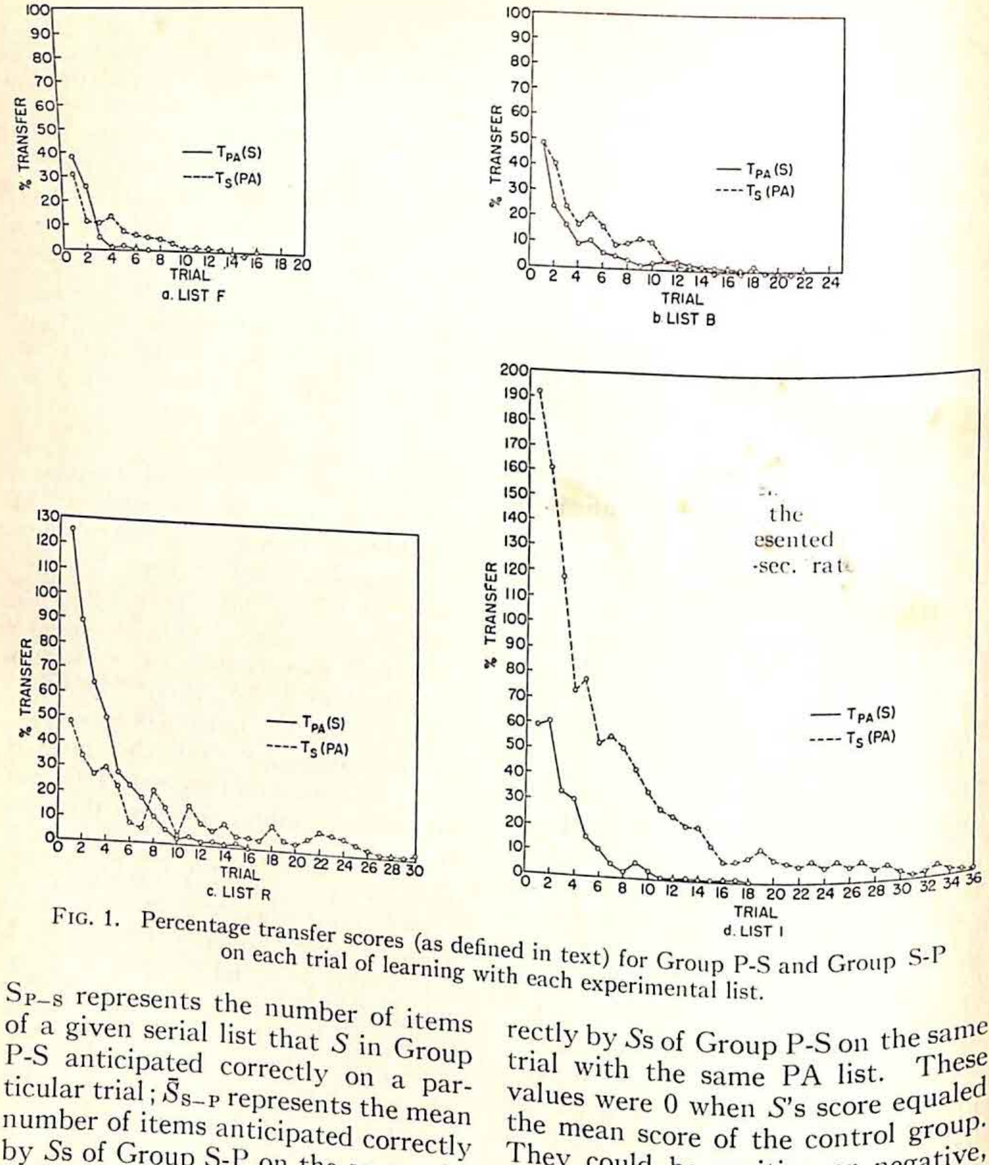


FIG. 1. Percentage transfer scores (as defined in text) for Group P-S and Group S-P on each trial of learning with each experimental list.

S_{P-S} represents the number of items of a given serial list that *S* in Group P-S anticipated correctly on a particular trial; $\bar{S}_{S-P}$ represents the mean number of items anticipated correctly by *S*s of Group S-P on the same trial with the same serial list. Likewise, PA transfer scores of Group S-P were computed by the following formula:

$$T_S(PA) = \frac{P_{S-P} - \bar{P}_{P-S}}{\bar{P}_{P-S}} \times 100.$$

P_{S-P} represents the number of items of a given PA list that *S* in Group S-P anticipated correctly on a particular trial; and $\bar{P}_{P-S}$ represents the mean number of items anticipated correctly by *S*s of Group P-S on the same trial with the same PA list. These values were 0 when *S*'s score equaled the mean score of the control group. They could be positive or negative, and their maximum value was limited only by the maximum possible score.

Figure 1 presents separate graphs for each experimental list, showing the mean transfer scores on each learning trial. The distribution of these scores was approximately normal, and a complete analysis of variance was performed on them for Trials 1–10. The mean differences between trials, between learning tasks, and between lists were all significant at $p < .001$.

TABLE 2

Analyses of Errors Made by Each Group throughout Learning

Errors	Lists							
	F		B		R		I	
	n	%	*n*	%	*n*	%	*n*	%
A. Serial Learning, Group S-P								
Within-list errors	48	36.09	128	38.44	165	28.06	206	39.84
Adjacent backward	0	0.00	0	0.00	1	0.17	1	0.19
Remote	48	36.09	128	38.44	164	27.89	205	39.65
Forward	44	33.08	109	32.73	135	22.96	196	37.91
Backward	4	3.01	19	5.71	29	4.93	9	1.74
Extralist errors	1	0.75	3	0.90	4	0.68	5	0.97
Failures to respond	84	63.16	202	60.66	419	71.26	306	59.19
Total	133		333		588		517	
B. Serial Learning, Group P-S								
Within-list errors	10	30.30	28	22.58	25	30.12	47	20.89
Adjacent backward	0	0.00	2	1.61	1	1.20	0	0.00
Remote	10	30.30	26	20.97	24	28.92	47	20.89
Forward	9	27.27	23	18.55	22	26.51	41	18.22
Backward	1	3.03	3	2.42	2	2.41	6	2.67
Extralist errors	0	0.00	0	0.00	0	0.00	0	0.00
Failures to respond	23	69.70	96	77.42	58	69.88	178	79.11
Total	33		124		83		225	
C. PA Learning, Group P-S								
Within-list errors	318	79.10	600	85.60	835	77.03	901	66.15
Adjacent backward	207	51.49	383	54.64	437	40.31	429	31.50
Remote	111	27.61	217	30.96	398	36.72	472	34.65
Extralist errors	1	0.25	5	0.71	0	0.00	4	0.29
Failures to respond	83	20.65	96	13.69	249	22.97	457	33.55
Total	402		701		1,084		1,362	
D. PA Learning, Group S-P								
Within-list errors	238	85.58	426	93.01	468	71.45	194	81.17
Adjacent backward	169	60.76	282	61.57	233	35.57	42	17.57
Remote	69	24.82	144	31.44	235	35.88	152	63.60
Extralist errors	0	0.00	5	1.09	1	0.15	1	0.42
Failures to respond	40	14.39	27	5.89	186	28.40	44	18.41
Total	278		458		655		239	

These *F*s were 112.30 ($df = 9/648$), 14.09 ($df = 1/72$), and 32.11 ($df = 3/72$), respectively. The interaction between lists and learning tasks was significant, F (3, 72) = 23.35, $p < .001$, and so was the interaction between lists and trials, F (27, 648) = 9.89, $p < .001$. Finally, the List × Trial × Task interaction was significant, F (27, 648) = 12.01, $p < .001$. No other sources of variation were significant.

Figure 1 shows that positive transfer occurred for both learning tasks with all lists. For Lists F and B the amount of transfer on the serial task about equaled that on the PA task. The most striking differences, however, occurred for List R and List I. The $T_{PA}(S)$ scores of List R greatly exceeded the $T_S(PA)$ scores, though this difference disappeared on later trials. For List I the $T_S(PA)$ scores greatly exceeded the $T_{PA}(S)$ scores.

Analysis of errors.—Table 2 presents the total number of errors made by each group throughout learning. These errors have been classified as "within-list errors," "extralist intrusions," or "failures to respond." A within-list error is defined as an appropriate response produced at the wrong time; such errors were further sorted into "adjacent backward associations" and "remote associations." Whenever Item n in serial learning elicited Item $n-1$ instead of Item $n+1$, this error was called an adjacent backward association; the category was also used when an item in PA learning elicited its stimulus instead of its response. Remote associations of serial learning were further classified as "remote forward" or "remote backward" associations. Extralist intrusions were so rare for all lists that they will not be discussed further.

The outstanding error of serial learning was the failure to respond. Failures were proportionately more frequent with every list during serial learning than during PA learning (all z's $\geqslant 7.47$, $p < .001$). Ratios of the number of failures in PA learning to those in serial learning (Table 2, Sections A and C) were 0.99 for List F, 0.48 for List B, 0.59 for List R, and 1.49 for List I. The corresponding ratios of total errors, however, were 3.02, 2.11, 1.84, and 2.63.

List R's percentage of failures in serial learning was particularly large, and this result probably accounts for its relatively small percentage of remote errors. Except for List R, the lists induced still higher percentages of failures in the serial task of Group P-S. The increase was statistically significant for List B ($z = 3.35$, $p < .001$) and for List I ($z = 5.24$, $p < .001$). Thus, one effect of the PA task on serial learning was to increase the percentage of failures to respond.

Backward associations hardly ever occurred during serial learning, and backward remote errors were also rare. Only .08 of the remote errors to List F were backward errors; to List B, .15; to List R, .18; and to List I, .04. Each of these proportions differed significantly from the corresponding proportion of forward errors (all z's $\geqslant 5.83$, $p < .001$). Thus, during serial learning S is not apt to produce backward errors (adjacent or remote). This result suggests that S monitors his responses, and rather than emit erroneous "backward" responses, he does not respond. At least in this sense, preceding items of the serial list must contribute to the effective stimulus to which S responds.

The outstanding error of PA learning was the adjacent backward error. With List F, .65 of the within-list errors were of this type; with List B, .64; with List R, .52; and with List I, .48 (Table 2, Section C). Backward errors were proportionately more frequent for Lists F and B ($z = 10.16$, $p < .001$), probably because adjacent items were strong associates. Although within-list errors as a whole were proportionately more frequent to List R than to List I, the constituent errors within the category were similarly proportioned ($p > .10$). After the serial task (Table 2, Section D) the proportion of backward errors in List I dropped strikingly ($z = 4.35$, $p < .001$), and the proportion of remote errors increased ($z = 8.46$, $p < .001$). Whatever produced this change must be due to the nature of the serial task for List I.

In the following analysis List F was the standard since its items are arranged optimally for the chaining hypothesis. The total number of errors on each list (Table 2, Section A) was divided by the total number of errors on List F. These ratios were 2.50 for List B, 3.89 for List I, and 4.42 for List R. If *S* learns a given list by the method he learns List F, his error pattern for that list should resemble the error pattern of List F. That is, ratios of error frequencies for individual categories should approximate the ratio computed for the total frequencies. For List B, these ratios were all nearly 2.50, the ratio for total errors; they ranged from 2.40 (failures) to 2.67 (remote). (Categories were ignored if they contained 10 errors or less.) The ratios for List I ranged from 3.64 (failures) to 4.45 (forward remote). For List R they were especially variable, ranging from 3.07 (forward remote) to 4.99 (failures). Thus, List R's error pattern deviates most from List F's, and the chaining hypothesis seems least valid for it.

A similar analysis was performed on the PA errors (Table 2, Section C). The ratio of total errors to those of List F was 1.74 for List B, 2.70 for List R, and 3.39 for List I. For individual error categories of List B, the ratios ranged from 1.16 (failures) to 1.95 (remote). For List R they ranged from 2.11 (adjacent backward) to 3.59 (remote). For List I they ranged from 2.07 (adjacent backward) to 5.51 (failures). Thus, when *S has* to learn associations between two words (PA learning), his error pattern on List R more nearly conforms to his pattern on Lists F and B. In serial learning, though, he does not have to learn such associations and his error pattern is more deviant.

Finally, the data were examined to learn which error categories were most affected by the transfer between tasks. Differences were computed between corresponding entries of Sections A and B, and each difference was expressed as a fraction of the entry in Section A. These "proportionate decreases" for total errors were: .86 for List R, .75 for List F, .63 for List B, and .56 for List I. For individual error categories of List R, they were uniformly high, ranging from .84 (forward remote) to .93 (backward remote). List I's decrease in forward remote errors was .79, but its decrease in failures was significantly smaller, .42 ($z = 8.44$, $p < .001$). Thus, the PA task reduced forward remote errors for List I but preserved the failures to respond.

Similar measures were computed from Sections C and D. The proportionate decreases in total errors were: .31 for List F, .35 for List B, .40 for List R, and .82 for List I. These figures largely reflect the decreases in adjacent backward errors: .18 for List F, .26 for List B, .47 for List R, and .90 for List I. The corresponding decreases in remote errors were .38, .34, .41, and .68; those in failures to respond were .52, .72, .25, and .90. Thus, List I benefitted from the serial task in several ways, but least of all ($z \geqslant 8.15$, $p < .001$) in remote errors. This result implies that remote associations in List I played a significant role during serial learning, a point that is further examined below.

DISCUSSION

These results suggest that the conclusion drawn by Primoff (1938) and Young

(1959) is of limited generality. Their result has been replicated with List R, which contained internal interference. In such lists the transfer from serial to PA learning seems to be reliably less than that from PA to serial learning. Thus, *S* can transfer to the serial task associations that he learned by the PA method; but when he learns the serial task first, his advantage in PA learning is less marked.

When adjacent items were deliberately selected to be strong associates (Lists F and B), the amount of transfer from serial to PA learning about equaled that from PA to serial learning. In the serial task of these lists, the effective stimulus for an item may plausibly be the single preceding item.

There was only one major difference between Lists F and B throughout the present study: List F was uniformly easier. This result has an implication about the role of backward associations in PA learning. To say that backward associations are formed during PA learning accurately describes the problem, but on two counts it fails to explain the difference between serial and PA learning. First, the difference between tasks was no greater for List F than for List B, even though both tasks with List B contained strong backward associations. Second, if backward associations are formed during PA learning, the percentage of backward errors should be particularly large for List B. Table 2, however, shows that Lists F and B did not differ in this respect.

When internal interference was minimized (List I), *S* did seem to form S-R associations between adjacent serial items since his PA performance was so facilitated. However, three results suggest that even for List I, the chaining hypothesis is oversimplified. First, the PA task required about twice as many trials as the serial task. Second, the amount of transfer for Group P-S was somewhat less than that of Group S-P. And third, remote errors persisted so in Group S-P's PA learning.

To account for these results, the present authors propose the following hypothesis. When a serial list is relatively free of internal interference (List I) or when the internal associations are organized (Lists F and B), the effective stimulus is not a single item preceding the response, but a *cluster* of preceding items. In the serial list A-B-C-D-E, for example, the cluster of items BCD might serve as the effective stimulus for Response E. (A cluster of any size would do, and the size of such clusters might even vary throughout the list.) Each element of the cluster would become associated with the response, and the nearer the element to the response, the greater the associative strength (McGeoch, 1936). Thus, Response E would be more strongly associated with D than with C; but, assuming that such associative strengths summate, the entire cluster BCD would elicit Response E more strongly than would any of the individual items. And in this way, remote associations play an important role in the serial task. (In List F, however, D so strongly elicits E that the association between B and E or between C and E would contribute relatively little to the total associative strength between the cluster and the response.)

The elements within the cluster BCD would also become associated with one another during serial learning. Element D, for example, would have been a response to the earlier cluster ABC and would therefore be associated with both B and C. The senior author has unpublished data on PA tasks where the stimuli were each clusters of elements. When the elements of a cluster were interassociated, performance was facilitated. This facilitation may be sufficiently great to account for the superiority of serial learning over PA learning.

Thus, when *S* learns List I serially, the cluster BCD elicits the response E; as a consequence, D elicits E more strongly than any other single item. If *S* is then asked to learn paired associates like D-E, this PA task would be facilitated. However, if *S* learned the PA task first, only the D-E association would have been formed. To learn the list serially, he would still have to form further associations before the cluster could elicit the response. Thus, the transfer from serial to PA learning would exceed that from PA to serial learning. With List F, on the other hand, the associative strength between D and E approximately equals the strength between the cluster and E, and the transfer would be about equal in both directions.

This analysis has a further implication. If *S* were responding to the stimulus cluster BCD, he would not produce any of its elements as a response, and backward errors would be minimized. Thus, this analysis could account for all the major results of the present study—the relative speed of serial learning, the absence of backward errors in serial learning, and the difference between the two directions of transfer.

Young (1962, Exp. II) has presented data which seem to disconfirm this hypothesis. After *S* mastered a serial list of adjectives, PA items were presented. These items were composed of three consecutive items of the serial list, Items $n - 2$, $n - 1$, and n. Items $n - 2$ and $n - 1$ were presented simultaneously as a stimulus cluster, and Item n was the response. Control pairs were also included in the task, and *S*'s performance on the experimental pairs was hindered.

However, this evidence does not disconfirm the proposed hypothesis for three reasons: (*a*) In an earlier experiment of the same study (Exp. I), Young examined the transfer from the serial task to Primoff's PA task. The transfer was negligible, and Young's list thus seems to resemble List R of the present study. The proposed hypothesis, however, only concerns Lists F, B, and I. (*b*) The associative relationship among Young's items is not known, and their internal association may resemble that of List R. (*c*) Because of Young's procedure, Exp. II does not directly test the hypothesis that a cluster of items is the effective serial stimulus. The two items of his stimulus clusters appeared simultaneously and thus failed to capture the temporal nature of a serial task. The present writers feel that for a more direct test, the elements of each cluster should appear singly as they do during the serial task. Then *S* could not instruct himself to ignore Item $n - 1$ during the PA task, as he could with Young's procedure.

REFERENCES

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312.

DEESE, J. From the isolated verbal unit to connected discourse. In C. Cofer (Ed.), *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

HOROWITZ, L. M. Free recall and ordering of trigrams. *J. exp. Psychol.*, 1961, **62**, 51–57.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MCGEOCH, J. A. The direction and extent of intraserial association at recall. *Amer. J. Psychol.*, 1936, **48**, 221–245.

MURDOCK, B. B. Backward associations in transfer and learning. *J. exp. Psychol.*, 1958, **55**, 111–114.

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

PRIMOFF, E. Backward and forward associations as an organizing act in serial and in paired-associate learning. *J. Psychol.*, 1938, **5**, 375–395.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota Norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* Teachers College, Columbia University, Bureau of Publications, 1944.

UMEMOTO, T., & HILGARD, E. R. Paired-associates learning as a function of similarity: Common stimulus and response items within the list. *J. exp. Psychol.*, 1961, **62**, 97–104.

YOUNG, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

YOUNG, R. K. Paired-associate learning when the same items occur as stimuli and responses. *J. exp. Psychol.*, 1961, **61**, 315–318.

YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, **63**, 307–313.

(Received May 15, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 362–367

REPETITION FACTORS IN ONE-TRIAL LEARNING OF PAIRED ASSOCIATES[1]

JAMES H. REYNOLDS[2]

Syracuse University

A 1-trial paired-associate task was used in a P-T1-T2 design to test 4 hypotheses inferred from current research on 1-trial learning and immediate memory. The results confirmed the 1st 2 hypotheses tested, indicating (a) that a correct association is always available immediately following presentation of simple paired-associate materials, and (b) that this immediate association is weak and decays rapidly. The hypotheses that (c) covert repetition and (d) overt repetition of immediate associations lengthen their recall availability were partially confirmed. An analysis of the effect of list position upon response availability indicated that the formation of a response set, as well as covert and overt repetition, facilitates recall of associations that are formed during a single presentation.

A strict all-or-none interpretation of the one-trial learning hypothesis assumes that repetition of material to be learned contributes to the acquisition process only insofar as it permits repeated opportunity for the one trial in which learning actually occurs to take place. Continued repetition beyond this crucial single trial, whether overt or covert, should play no further role in the learning process. However, the results of several recent studies on one-trial verbal learning have given evidence contrary to this all-or-none interpretation. Rock (1957) and Rock and Heimer (1959) have concluded that although initial learning takes place on a single trial, repetition subsequent to the single acquisition trial strengthens the association that has been acquired, making the association more resistant to forgetting. Estes, Hopkins, and Crothers (1960) have reached a similar conclusion, apparently qualifying Estes' (1960) earlier assumption that repetition following original learning has no effect upon the strength of the single-trial association.

These restrictions on the one-trial learning concept suggest that, although acquisition may indeed occur on a single trial in the sense that what is learned becomes observable all at once, the total process of learning is not an all-or-none affair. Alternatively, the experimental results reported by Rock and Estes and their associates are consistent with investigations of immediate memory recently made by Peterson and Peterson (1959) and Peterson, Saltzman, Hillner, and Land (1962). These latter studies have indicated that the probability of recall immediately following a presentation is close to 1.0, that this short-term retention diminishes rapidly in time, but that repetition during this interval facilitates later recall. Taken together, their results imply that the recent evidence for one-trial verbal learning may depend upon two factors; the occurrence of a momentary after-sensation that follows stimulus presentation, and the extent to which repetition takes place during the short-term retention period.

---

[1] This paper is based on a dissertation submitted to the Graduate School of Syracuse University in partial fulfillment of the requirements for the PhD degree. The author is indebted to W. R. McAllister for his many helpful suggestions and guidance during the course of the study.

[2] Now at the University of Pittsburgh.

The following four hypotheses, based upon the above implication, were tested in a paired-associate design that has previously been used in one-trial learning experimentation: (*a*) immediately following presentation of an S-R pair the probability of recall is 1.0 for that pair; (*b*) a rest interval following presentation of an S-R pair permits covert repetition during the short-term retention period, having the effect of lengthening the time the immediate association is retained; (*c*) single-trial associations made during presentation of a series of paired associates are weak, and decay rapidly with time when opportunity for repetition is minimal; and (*d*) overt repetition of associations that are formed on a single trial lengthen the time in which they remain available.

METHOD

Subjects.—A total of 120 undergraduate college students, 66 men and 54 women enrolled in an introductory psychology course, participated in the study as part of a course requirement.

Apparatus and materials.—Using a paired-associate task, variations of the Presentation-Test 1-(Rest)-Test 2 (P-T1-T2) paradigm employed by Estes (1960) were administered under four conditions to test the hypotheses. For the P trial a Patterson memory drum, Model S-1, was used to present eight number-letter pairs to each group in the following order: 74-J, 38-C, 41-V, 93-S, 85-R, 29-X, 57-F, 62-W. The stimulus elements (numbers) were chosen so that no two pairs had either a first or a second digit alike, and consonant letters were chosen so that no two response elements were the same. The memory drum was also used to present the stimulus elements alone during T1 and T2 trials. Different list orders for the P trial and both T trials were preselected on a random basis, and these fixed orders were used for all groups. The rest intervals between learning and test trials were timed with a stop watch.

A questionnaire was administered as an interpolated task to all groups during the rest period. The questionnaire, assumed to be of low interference value, required *S*s to judge the ambiguity of a series of sentences concerning educational objectives, and to write a short essay expressing personal opinions about these objectives. Another interpolated task was a word list used for a word association test, administered to one of the groups as a substitute for the T1 trial. This list contained the following eight words: BRAVE, WISH, SHY, BRIGHT, HAPPY, MOTHER, RUSH, WATCH.

Design and procedure.—Men and women *S*s were assigned randomly to one of the four groups as they appeared for the experiment. An Immediate Memory group (Group IM), consisting of 13 women and 15 men, was presented with the P and T1 materials in an order designed to test the first hypothesis. Each *S* was presented with an S-R pair for 4 sec., followed by a 4-sec. rest interval; then the stimulus member alone for that pair was presented for 4 sec., followed by another 4-sec. rest interval during which *S* was required to give the response member verbally if possible. This sequence was repeated until all eight pairs had been presented and immediately tested. Following this combined P-T1 period, *S* began a 5-min. interpolated rest period in which the questionnaire was administered to prevent rehearsal. After the rest period the T2 trial was presented, in which stimulus elements alone were shown on the drum for 4 sec. each, interspersed with 4-sec. rest intervals during which *S* was required to give the response member verbally if it could be recalled.

Three other groups, one consisting of two subgroups, received varied opportunities for either covert repetition (Co) or overt repetition (Ov) of the original list of pairs presented during the P trial. These groups were used to test the remaining three hypotheses.

A Co-Ov group, 13 women and 15 men, was given the P trial, followed by T1, the questionnaire, and the T2 trial, an order identical to that used by Estes (1960). For this group, 4-sec. interpair rest intervals during the P trial permitted time for covert repetition of the pairs presented. A 30-sec. interval following the P trial was used to give instructions for T1. The T1 trial, during which *S* recalled aloud the response members, provided opportunity for overt repetition of the material learned. The questionnaire period and T2 trial that followed were identical to those used for Group IM.

Group N-Ov, consisting of 13 women and 15 men, was exposed to the same procedural order as Group Co-Ov. The T1 trial permitted opportunity for overt repetition of pairs recalled from the P trial. However, opportunity for covert repetition during the P trial was limited by the omission of the 4-sec. interpair rest intervals. According to the second hypothesis, this limitation should have a negative effect upon recall during T1 and T2.

In Subgroups Co-N and Co-WA, covert repetition was possible during the P trial, in which 4-sec. rest intervals occurred between presentations of the number-letter pairs. However, the T1 trial was eliminated for both subgroups to prevent overt repetition. Subgroup Co-N, consisting of 7 women and 10 men, received a P trial identical to that of Group Co-Ov. Immediately following the P trial, these *S*s began a $6\frac{1}{2}$-min. rest interval during which they filled in the questionnaire (the extra $1\frac{1}{2}$ min. being the time during which other groups were exposed to T1 tasks). The test trial following the questionnaire was identical to the T2 trials received by Group Co-Ov and the other groups. Subgroup Co-WA, 8 women and 11 men, also received the same P and T2 trials as Group Co-Ov, but 30 sec. after the P trial these *S*s were presented with the eight-item word association (WA) task on the memory drum. This task was a control for the T1 task administered to the Co-Ov group, giving Subgroup Co-WA *S*s the same attention and response conditions as Group Co-Ov at the time of T1 (i.e., looking at the drum, responding verbally to the drum stimuli), but offering no opportunity for overt repetition of the associations made during the P trial.

The third hypothesis was tested by comparing the final test performance of each of these subgroups with the T1 performance of the Co-Ov group, and the fourth hypothesis was tested by comparing performance on the test trials of subgroups with the T2 performance of the Co-Ov group. The response measure employed in all tests of hypotheses was the number of correct responses made by the groups during T1 and T2.

The *E* sat behind the memory drum, out of *S*'s view at all times except while giving instructions. After the last test trial, the last 90 of the 120 *S*s were asked, "What method or system did you use to learn the pairs when they were first shown to you?" The *E* recorded whether *S* reported covert repetition, the use of mnemonic aids, or some other method.

Instructions.—To avoid awareness that T2 would follow the interpolated task, all *S*s were told that they would participate in two experiments, the first a brief verbal learning task and the second a polling of opinion concerning educational objectives. All *S*s were then told that in the first experiment a series of number-letter pairs would appear in the memory drum window, and their task was to learn the number and the letter that was paired with it. Group IM *S*s were told in addition that between each pair presented the number of *any one* of the pairs they had seen previously would appear in the window, and that whenever just a number appeared they were to say aloud the letter that had been paired with it. The instruction that *any* previous number might appear was given to insure that the IM *S*s would try to learn each pair rather than just the response member, which was really all that was necessary for their T1 task.

In the 30 sec. following the P trial, the groups were given varying instructions. Groups Co-Ov and N-Ov were told that they would next see just the numbers of the pairs in the window and their task would be to say aloud the letter that had been paired with the number shown. The *S*s in Group Co-WA were told that a series of words would appear in the drum window one at a time, and they were to say aloud the first word they associated with each word as it appeared. The *S*s in Group Co-N were given instructions about filling in the questionnaire.

After *S*s finished T1 (or the word association task for Group Co-WA), they were told they would next participate in the second experiment. The questionnaire was described to them, and they were instructed to begin working, seated at a table that was placed beside the memory drum at a right angle to it. At the end of the 5-min. period ($6\frac{1}{2}$ min. for Group Co-N), they were told to stop working on the questionnaire, to face the memory drum again, and to say aloud the letters that had been paired with the numbers they were about to see. On all tests, they were instructed to guess if they were not sure of the correct answer.

Before being dismissed from the experiment, *S*s were asked not to tell other students the nature of the tasks they had just completed.

RESULTS

Table 1 presents the mean number of correct responses and the standard deviation for each group on T1 and T2. Since preliminary *t* tests of sex differences within each group in mean number of correct responses on T1 and T2 yielded insignificant values, the results for both men and women were pooled. A *t* test of the difference between the means of Subgroups Co-N and Co-WA on their single test trial was significant, $t\ (34) = 2.58$, $p < .02$,

TABLE 1

MEANS AND *SD*s OF CORRECT RESPONSES FOR EACH GROUP ON TEST 1 AND TEST 2

		Group				
		IM	Co-Ov	N-Ov	Co-N	Co-WA
Test 1	Mean	7.57	2.57	1.86	—	—
	SD	.84	1.32	1.39	—	—
Test 2	Mean	2.75	2.57	1.43	1.17	2.16
	SD	1.86	1.57	1.14	1.01	1.26

indicating that they should be treated separately.

Although distributions of the number of correct responses were somewhat skewed, they were similar in shape among all of the groups, and differences in variance among the groups were within the .05 limits of chance. Consequently, *t* tests rather than nonparametric tests were used in making the comparisons necessary to test the hypotheses (Anderson, 1961; Boneau, 1960). The resulting *t* ratios are presented in Table 2.

Hypothesis a.—Group IM recall on the test trial given 4 sec. after initial presentation was 94%. Twenty-one of the 28 *S*s responded correctly to all eight stimuli, and only 1 *S* missed as many as three. With a few exceptions, probably attributable to momentary inattention or to the 4-sec. lapse between P and T1, the results support the hypothesis that recall probability immediately following a single presentation is 1.0.

TABLE 2

t VALUES OBTAINED FROM COMPARISONS OF GROUP MEANS TESTING HYPOTHESES *b*, *c*, AND *d*

Hypothesis Tested	Groups Compared	*df*	*t*
b	Co-Ov (T1)–N-Ov (T1) Co-Ov (T2)–N-Ov (T2)	54 54	1.97* 3.11**
c	Co-Ov (T1)–Co-N Co-Ov (T1)–Co-WA	43 45	4.02** 1.08
d	Co-Ov (T2)–Co-N Co-Ov (T2)-Co-WA	43 45	2.07* .95

* $p \leq .05$.
** $p \leq .01$.

Hypothesis b.—The hypothesis that interpair rest intervals during the P trial lengthens the time that the association is retained because of increased opportunity for covert repetition was tested by comparing Group Co-Ov recall with that of Group N-Ov. The significant mean differences between these groups in both T1 and T2, as shown in Table 2, were in the direction predicted; i.e., the 4-sec. interpair interval received by Group Co-Ov facilitated recall of associations that were made during a single presentation. Of those *S*s asked, the majority in both the Co-Ov and the N-Ov groups reported rehearsal as either their only or their major method of learning, and some *S*s in both groups complained that after the first few pairs they were unable to rehearse all of the list between pair presentations. While not conclusive, these reports suggest that the significant differences between the groups were due, at least in part, to the difference in amount of covert repetition that was permitted to take place during the P trial.

Hypothesis c.—The hypothesis that single-trial associations are weak and subject to rapid decay was tested by comparing the results of the T1 given to Group Co-Ov immediately after the P trial with the final and only test given to Subgroups Co-N and Co-WA, which came after a longer interval. The difference in mean number of correct responses between Group Co-Ov (T1) and Subgroup Co-N is significant beyond the .05 level in the predicted direction. As shown in Table 2, however, the difference in the means between Group Co-Ov (T1) and Subgroup Co-WA falls short of the .05 level of significance. Thus the third hypothesis was only partially confirmed.

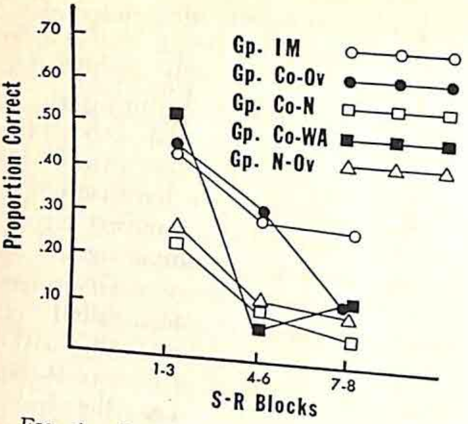


FIG. 1. Proportion correct on final test trial for all groups on blocks of S-R pairs representing the beginning, middle, and end of the original list.

Hypothesis d.—According to the fourth hypothesis, the recall of Group Co-Ov at the time of T2 should be superior to both subgroups because of its opportunity for overt repetition of items during T1. As Table 2 indicates, the predicted difference between the mean number of correct responses for Group Co-Ov (T2) and Subgroup Co-N is significant at the .05 level. The difference between the Group Co-Ov (T2) and Subgroup Co-WA means is well within chance limits, however, demonstrating again the unexpected effect of the WA task in facilitating the retention of the original associations.

List position.—To analyze further the unpredicted effect of the WA task, the mean proportions correct on T2 for all groups on blocks of items representing the beginning, middle, and end of the original list were plotted. As may be seen in Fig. 1, the effect of list position upon retention was generally similar for Groups Co-N and N-Ov. The other groups, each receiving a different T1 treatment, show varying degrees in the extent to which list position affected retention. Failure to satisfy *t* test assumptions prevented comparisons of the mean proportions of Groups IM, Co-Ov, and Co-WA in the various list positions. However, fourfold contingency tables testing differences in the number of *S*s giving one or more correct responses revealed that retention differences between Co-WA and both the IM and Co-Ov groups were significant for items in the middle list position (for $df = 1$, $\chi^2 = 5.16$, 9.85, $p < .05$, $.01$, respectively). No significant differences among the three groups were found for items in the beginning or end positions. Although the IM, Co-Ov, and Co-WA treatments all facilitated overall retention, they apparently had differential effects upon recall of the middle list items.

DISCUSSION

The 94% recall of Group IM 4 sec. following a single presentation supports the finding of Peterson, Saltzman, Hillner, and Land (1962) that recall probability immediately following presentation is close to 1.0. Peterson and his associates also showed that the nearly 1.0 probability decreases rapidly, over periods as short as 16 sec., when opportunity for rehearsal is minimal. The low T1 recall proportions of the Co-Ov and N-Ov groups in the present study indicate again that the 1.0 probability diminishes in a short time. However, the superiority of Group Co-Ov over Group N-Ov on T1 suggests that the rapidity of the decrement in recall availability following a single presentation is significantly changed when opportunity for covert repetition during the high probability period is provided. This effect of covert repetition upon retention during the T1 trials, which occurred at least 30 sec. after the last *P* trial, is an extension of Peterson and Peterson's (1959) finding that either overt or covert repetition facilitates retention of short-term association up to periods of 18 sec.

Comparisons of the T2 performances of the Co-Ov, Co-N, and N-Ov groups offer evidence of relationships among covert and overt repetition, and the time in which each takes place, that affect long-term retention of single-trial associations in ways not demonstrated previously. In an investigation of one-trial learning, Lockhead (1961) has reported evidence that opportunity for rehearsal immediately following a presentation raises recall probability at a later time. This finding agrees with the hypothesis that evidence for one-trial learning depends upon repetition of immediate memory associations. However, the equivalent T2 recall of the Co-N and N-Ov groups indicates that opportunity for immediate rehearsal, if taken alone, is no more effective in facilitating longer term retention than is overt repetition at a time when immediate memory associations have presumably disappeared. Alternatively the IM and Co-Ov groups, each receiving opportunity for both covert and overt repetition but at varying times, showed consistently greater retention than either Co-N or N-Ov. Apparently the optimal condition for retention of short-term associations formed on a single trial is one which permits overt repetition as well as immediate covert repetition.

The varying recall performances of the IM, Co-Ov, and Co-WA groups imply that overt repetition following immediate covert repetition makes a dual contribution to raising recall availability at a later time. The facilitating effect of the WA task upon retention, which bears some similarity to increased retention effects found in studies of warm up (Adams, 1961), suggests that the contribution of overt repetition to increasing recall availability is simply the establishment of a response set. But the significant T2 recall differences between Co-WA and both the IM and the Co-Ov groups in the middle list position indicate that actual overt repetition facilitated retention of certain items, traditionally considered the most difficult to learn, whose recall availability was not increased by response practice alone. The implication is that overt repetition has a strengthening effect upon existing associations as well as offering response practice, although the strengthening effect is difficult to observe because of the major contribution that response practice makes to recall facilitation.

Taken together, the data imply that immediate memory associations are almost invariably present following a single presentation of simple learning materials, and that the combination of several factors occurring during or soon after this short-term retention period contributes to the recall availability of the immediate associations at a later time. These findings support a conclusion that immediate memory and the reported evidence for one-trial learning are not independent phenomena but temporal segments of a continuous associative process.

REFERENCES

Adams, J. A. The second facet of forgetting: A review of warm-up decrement. *Psychol. Bull.*, 1961, **58**, 257–273.

Anderson, N. H. Scales and Statistics: Parametric and nonparametric. *Psychol. Bull.*, 1961, **58**, 305–316.

Boneau, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, **57**, 49–64.

Estes, W. K. Learning and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

Lockhead, G. R. A re-evaluation of evidence of one-trial associative learning. *Amer. J. Psychol.*, 1961, **74**, 590–595.

Peterson, L. R., & Peterson, J. M. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58**, 193–198.

Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Rock, I., & Heimer, W. Further evidence for one-trial associative learning. *Amer. J. Psychol.*, 1959, **72**, 1–16.

(Received May 17, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 368–376

AN INVESTIGATION OF RECODING IN FREE RECALL[1]

BURTON H. COHEN

Lafayette College

4 studies were conducted to evaluate G. A. Miller's chunk hypothesis. 70-word lists comprising 10 exhaustive and 10 nonexhaustive categories of words were presented for free recall. The results indicated that: (a) whether category recall was inferred from word recall or measured directly, between 10 to 14 categories were represented in recall; (b) significantly more words of exhaustive categories were recalled; (c) the 2 category types were used equally in recall; (d) there was a strong tendency to cluster related words in recall; and (e) an index of interitem associative strength, derived from word association norms, discriminated differences in recall between the 2 category types and was significantly related to within-category recall. The results were interpreted as providing support for Miller's position regarding recoding.

Miller (1956b) presented the hypothesis that large amounts of elementary input information can be handled by a limited span of immediate memory if the input information is recoded into chunks. While Miller is not specific, he tentatively identifies the chunk as a "new name" (p. 93) for a group of input items while at another point he states that chunks are "groups of words that go together" (p. 95). The purpose of the present paper is the attempt to extend the chunk hypothesis to the free recall of categorized word lists. Specific interest was focused on (*a*) the number of chunks processed in free recall and (*b*) the comparison of two types of chunks.

Lists of 70 words belonging to 20 categories were presented to *S*s for free recall. Half the categories were termed nonexhaustive (NE) since the words only partially exhausted the pool of words belonging to each category. For example, DOG, LION, HORSE, and BEAR are members of a category which may be labeled Animals. However, these four members represent only a small sample from a large pool of animal names. The other half of the categories were termed exhaustive (E) since the words exhausted or nearly exhausted the entire pool of words belonging to the category. For example, the words NORTH, SOUTH, EAST, and WEST are the only members of a category which may be termed Cardinal Points of the Compass.

Four experiments will be reported here. In Exp. I, the words were presented to groups of *S*s who wrote the words they could recall. In Exp. II, the words were presented to individual *S*s and recall was oral rather than written. For Exp. III, *S*s were tested in a group situation but each *S* exposed his own list and randomization by use of individual exposure devices. The *S*s in Exp. IV exposed the words of the lists by means of the individual exposure devices but were asked to recall the names (category labels) for the groups of words of the list. In addition to these experiments, word associations were obtained to the words of both lists.

---

[1] This study was supported by a research grant, M-4430(A), from the Public Health Service, the National Institutes of Health. The author is indebted to Peter A. Hut who assisted in the collection and analysis of the data of Exp. II.

METHOD

Two lists of 70 words each were constructed. Each list comprised 10 E and 10 NE categories of words. For each type of category there were 5 categories of 4 words and 5 categories of 3 words each. An attempt was made to equate intercategory and interlist word frequencies as determined from the Thorndike-Lorge (1944) G counts. Since the number of E categories comprising 3 and 4 words was quite small, a severe restriction was placed on the selection of the E categories and their frequency matched NE counterparts. Given this restriction, the 20 categories comprising each list were selected to minimize any obvious intercategory associations.

The NE categories and words were selected from a set of norms collected by Cohen, Bousfield, and Whitmarsh (1957). The E categories and words were chosen from a set which the author has accumulated. For the NE categories, the total pool of words for each category may be estimated from the Cohen, Bousfield, and Whitmarsh norms. Since these norms contain a number of irrelevant responses, a count was made of only those responses which were members of the category. The ratio of the three or four selected words to this total count was computed for each of the 20 NE categories. The mean ratio of exhaustiveness was .060 and the range was .032 to .098. While no claim can be made that all the E categories were completely exhaustive, the mean ratio of exhaustiveness is probably as high as one can achieve for these categories. Thus, the difference in the mean ratio of exhaustiveness for the E and NE categories of each list is probably close to the maximum which can be attained with 20 category lists. The words of each list, along with their Thorndike-Lorge frequencies are shown in Table 1. These two lists were used in all four experiments. To reduce cues which would favor the recall of one word or category over another, the lists were systematically randomized so that no word of a category ever followed a word from the same category.

Experiment I.—The words were typed in capital letters on Radio-Mats which were made into 2 × 2 in. slides. The slides were projected with a Kodak Cavalcade projector, Model 540. An automatic timing cycle projected each word for 3 sec. An additional second was required to change from one slide to the next. Three randomizations of each list were used. The *S*s were given blank sheets of paper on which to write their recall.

Six groups of male undergraduate students at Lafayette College were randomly assigned to each of the three randomizations for each of the two lists. The groups varied in size from 18 to 40 *S*s. All told, 93 *S*s were presented with List 1 and 83 *S*s with List 2.

Experiment II.—The words were presented in exactly the same fashion as for Exp. I and the same three randomizations of each list were used. The *S*s were 48 male undergraduate students at Lafayette College who were randomly assigned to each of the two lists with the restriction that 8 *S*s receive each of the three randomizations of a list. The *S*s were tested individually in a moderately sound-proofed room. They were instructed to use an oral recall and were given to understand that *E* would write down their recall. A tape recorder was hidden in a closet of the testing room and a microphone was hidden behind a large speaker in the room. Just before the last words of the list were projected, *E* surreptitiously activated a mercury switch which turned on the tape recorder. In cases where the initial burst of recalled words was too fast to be accurately recorded by *E*, the tape playback was utilized to obtain the recall.

Experiment III.—The *S*s were tested in a group situation but each *S* was provided with an individual, manually operated, exposure device constructed from a small cardborad box. The words, typed triple spaced on adding machine tape, were exposed in a slot cut in the top of the box by turning a wooden dowel. The initial portion of all tapes contained a code number indicating the list and randomization number of the tape. A series of 10, three-digit numbers was used to provide practice in turning the dowel in a smooth and consistent fashion. The end of the practice list was indicated by a series of asterisks. A series of the numerals 1, 2, and 3 served as a warning device preceding the first word of the list proper. A 400-cps tone of approximately 200 msec. duration was recorded at 3-sec. intervals on tape. The *S*s practiced exposing the three-digit numbers in time with this recorded signal. They were then instructed to turn the dowel until the series of 3's appeared in the exposure slot. At this point, the specific instructions for the experiment were given. The tape recorder was then turned on and *S*s proceeded to expose the words of the list in time with the recorded signal.

The *S*s comprised two groups of male undergraduate students at Lafayette College. Each group contained 44 *S*s. Since both lists were used simultaneously, the exposure devices were distributed so that alternately

TABLE 1

Word Lists Used in the Experiments

List 1				List 2			
Exhaustive		Nonexhaustive		Exhaustive		Nonexhaustive	
Word	*f*	Word	*f*	Word	*f*	Word	*f*
ARMY	AA	COTTON	AA	SOLDIER	AA	RAIN	AA
NAVY	49	LINEN	47	SAILOR	A	SNOW	AA
MARINE	17	SATIN	15	MARINE	17	FOG	18
ANIMAL	AA	CHAIR	AA	BREAKFAST	A	HAT	AA
VEGETABLE	A	DESK	A	DINNER	AA	COAT	AA
MINERAL	38	COUCH	28	LUNCH	39	SHIRT	47
KNIFE	A	ROSE	AA	CIGARETTE	22	TRUCK	23
FORK	31	VIOLET	34	CIGAR	16	TAXI	17
SPOON	33	DAISY	38	PIPE	A	PLANE	A
BLONDE	17	TREASON	15	CAUCASIAN	1	ONYX	1
BRUNETTE	2	BURGLARY	1	ORIENTAL	14	RUBY	10
REDHEAD	—	LARCENY	1	NEGRO	47	PEARL	48
MASCULINE	8	VIOLIN	11	VOLT	2	WALTZ	3
FEMININE	8	FLUTE	9	AMPERE	—	TANGO	—
NEUTER	1	SAXOPHONE	1	OHM	—	POLKA	—
FRESHMAN	3	SARDINE	2	COCAIN	1	ASPEN	1
SOPHOMORE	3	MINNOW	3	HEROIN	—	SEQUOIA	1
JUNIOR	31	PERCH	23	MORPHINE	1	HEMLOCK	7
SENIOR	10	SALMON	14	OPIUM	7	CYPRESS	6
NORTH	AA	MOUNTAIN	AA	MOTHER	AA	DOG	AA
SOUTH	AA	HILL	AA	FATHER	AA	HORSE	AA
EAST	AA	VALLEY	AA	SISTER	AA	BEAR	AA
WEST	AA	RIVER	AA	BROTHER	AA	LION	A
ACE	3	BUNGALOW	8	INCH	AA	ROOF	AA
KING	AA	HOUSE	AA	FOOT	AA	WINDOW	AA
QUEEN	AA	HOTEL	A	YARD	AA	DOOR	AA
JACK	AA	CABIN	A	MILE	AA	FLOOR	AA
SUMMER	AA	ARM	AA	ADDITION	A	RUSSIA	A
WINTER	AA	LEG	AA	SUBTRACTION	1	BURMA	1
SPRING	AA	HEAD	AA	MULTIPLICATION	5	PERU	7
FALL	AA	FOOT	AA	DIVISION	A	SPAIN	A
PENNY	38	BISHOP	40	BASS	7	BAYONET	9
NICKEL	11	REVEREND	11	TENOR	6	DAGGER	8
DIME	11	PASTOR	11	ALTO	1	CUTLASS	2
QUARTER	AA	MINISTER	A	SOPRANO	1	STILETTO	—

Note.—*f* indicates frequency of occurrence from Thorndike-Lorge G count. Em dash (—) indicates word does not appear in G count. Categories are not given verbal labels since, in the main, such labels would be quite arbitrary.

seated *S*s received the same list. Twenty different randomizations of each list were drawn in accordance with the restriction mentioned earlier along with the additional restriction that every randomization start and end with a word from a different one of the 20 categories. Repetitions of the original 20 randomizations were used to accommodate the 44 *S*s tested on each list. As in Exp. I, *S*s were provided with blank sheets of paper on which to write their recall.

Experiment IV.—The procedure was identical with that of Exp. III except for the instructions to *S*s. In Exp. IV, *S*s were given instructions for exposing the words and were informed that further instructions would be given after the words had been exposed. These latter instructions requested *S*s to recall the names of the categories or groups of words rather than the words themselves. The *S*s were 39 male undergraduate Lafayette College students, 19 of whom received List 1 while 20 received List 2.

Instructions to Ss.—For Exp. I, II, and III, the instructions to *S*s were identical. The *S*s were told they would see a long list of common English words. They were told to watch the words as they were exposed. After the last word was exposed, they were to write (say) all the words of the list they could recall, in any order that occurred to them. The *S*s were informed they would have 5 min. for recall.

For Exp. IV, *S*s were given specific recall instructions after the words had been exposed. The *E* informed the *S*s that they had probably noted, in exposing the words, that the words could be classified into groups or categories of related words. The *S*s were instructed to write a label, name, or phrase for each of the groups or categories of words they could remember. The *E* emphasized to *S*s that they were to write the category names and not the individual words of the list. The *S*s were further instructed to use only the left side of the data sheet and were given 5 min. for the recall of the category names. After this recall period, *S*s were instructed to use the right side of the data sheet and write all the words they remembered that belonged to the categories they had listed. Again, 5 min. were allowed for this recall.

Word association data.—Because of the limited number of E categories comprising three and four words, no precise control over differences between the E and NE categories, other than the Thorndike-Lorge frequency matching, was feasible in constructing the lists. Deese (1959a), however, has demonstrated that an index of interitem associative strength derived from word association norms can be used to predict differences in the recall of groups of related words. In view of this, word association data were collected for the words of both lists.

Each *S* was given one of the exposure devices described in Exp. III. Each exposure device contained a tape with all 70 words from one of the two lists. After receiving standard word association instructions, *S*s exposed one word at a time and recorded their association to that word before exposing the next word. No more than 10 min. was required to finish the task. Aside from the longer time interval involved, this procedure was identical with the general technique by which the words were presented to *S*s in the experimental studies. Three different random orders of the 70 words of each list were used. A total of 54 male undergraduate Lehigh University students were randomly assigned to each list with the restriction that 18 *S*s receive each randomization. The norms are, therefore, based on a total of 54 *S*s per word. Lehigh University students rather than Lafayette College students were used since a large portion of the student body at Lafayette College had been exposed to the words of the lists at one time or another. Since both Lafayette and Lehigh are all-male schools with highly similar selection and admission procedures, no restriction on the generality of the normative data was anticipated. After tabulating the responses, the index of interitem associative strength, as described by Deese (1959a), was derived for each category of words.

RESULTS

Experiments I, II, and III.—The recall protocols were scored for two major response measures: (*a*) the number of words recalled and (*b*) the number of categories represented by the recalled words. For this latter measure, *S*s received credit for having represented the category in recall if at least one word of the category was recalled. Both these response measures were further decomposed into an E and NE component. The means and *SD*s for these response measures for both lists and for all three experiments are shown in Table 2. An analysis of variance of the total recall scores for both lists and all three experiments was computed. Since the result indicated a significant difference in the recall means, F (5, 306) = 9.55, $p < .001$, further analyses of the data will be reported separately for the three experiments unless otherwise noted.

Analyses of variance tests of the word recall scores showed that for Exp. I and III there was a significantly greater word recall for List 1 than List 2, F (1, 174) = 39.5, $p < .01$ and F (1, 86) = 5.3, $p < .05$, respectively. Experiment II showed

TABLE 2

Means and *SD*s of Words Recalled and Numbers of Categories Represented in Recall

Exp.	List		Words			Categories		
			Exhaustive	Nonexhaustive	Total	Exhaustive	Nonexhaustive	Total
I	1	Mean	21.22	16.58	37.78	7.03	6.86	13.89
		SD	5.01	4.88	8.05	1.29	1.66	2.26
	2	Mean	17.40	12.98	30.37	6.71	6.35	13.06
		SD	4.90	4.12	7.53	1.53	1.53	2.10
II	1	Mean	18.67	14.46	33.13	6.71	6.75	13.46
		SD	5.52	4.33	8.37	1.52	1.70	2.19
	2	Mean	18.33	13.71	32.04	7.33	6.75	14.08
		SD	5.83	5.71	10.94	1.55	1.80	3.01
III	1	Mean	19.05	14.70	33.75	6.43	6.39	12.82
		SD	4.49	5.13	7.84	1.28	1.83	2.47
	2	Mean	17.55	12.50	30.05	6.48	5.91	12.39
		SD	4.41	3.93	7.21	1.21	1.49	2.18

the same trend though the result was not significant. For all three experiments, the analyses of variance tests showed a significantly greater recall of words of E categories ($p < .01$ in every case). The interaction term was nonsignificant in every case.

Regarding the number of categories represented in recall, the analyses of variance tests detected only one significant difference—that between the two lists for Exp. I ($p < .05$). As can be seen from Table 2, however, this statistically significant difference reflects an increase of only 6.4% in the number of categories represented in recall in favor of List 1. Though the differences between the number of E and NE categories represented in recall were nonsignificant in all cases, a very slight trend is indicated by the fact that five of the six comparisons showed that more E categories were represented in recall. As a whole the analyses of word and category recall scores indicated that *S*s processed NE categories just as well as E categories even though word recall for the NE categories was significantly poorer.

The efficiency of recall of words of E categories may be demonstrated in the following manner. The total recall for a given list and experiment was partitioned into four additive components as follows: E, four-word categories; E, three-word categories; NE, four-word categories; and NE, three-word categories. A tabulation was made of the sum of the frequencies with which *S*s recalled one, two, three, or four words for the five categories of each of these components. (Note that this analysis is not concerned with the order in which the words were recalled.) From this tabulation, the percentage of the total component recall comprising occurrences of one, two, three, or four words of a category was determined. The results are shown in Fig. 1. The discrepancy between the E and NE categories is quite striking. Over both lists and all experiments, well over 50% of the total recall of words of E categories represented the complete recall of all

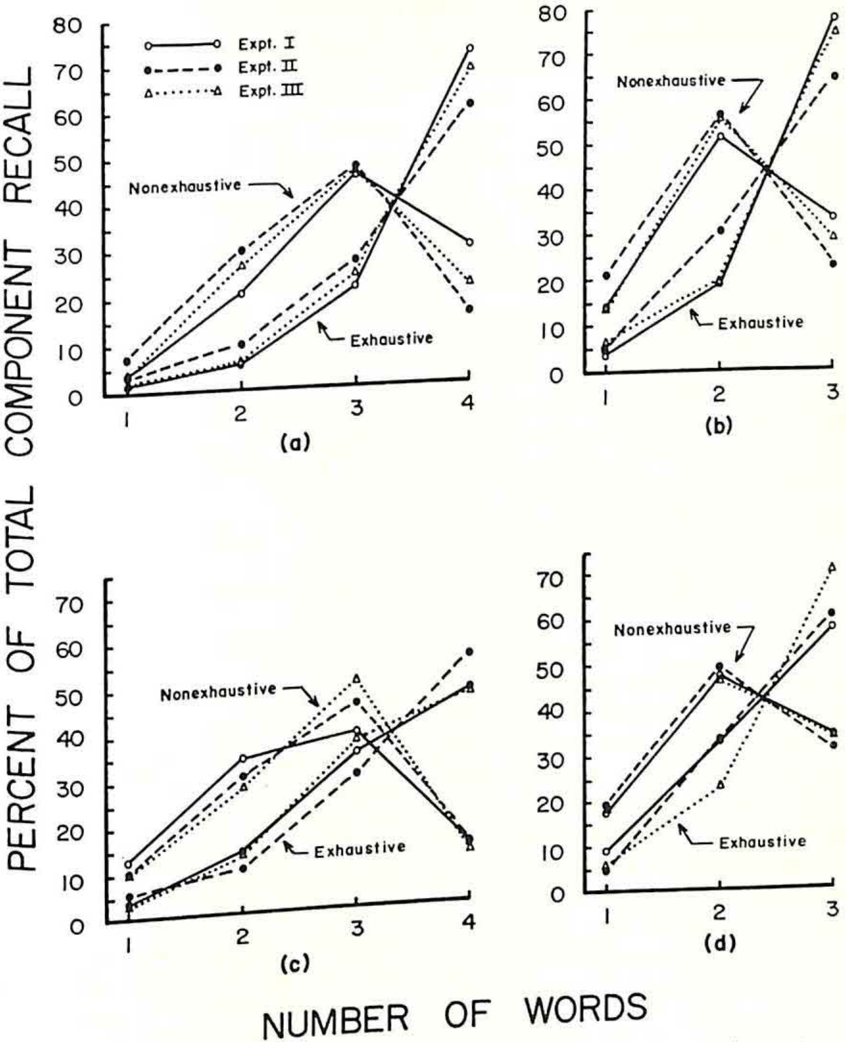


FIG. 1. Percentage of total component recall comprising occurrences of one, two, three, or four words for the five categories in each component. (List 1, four- and three-word categories are shown by 1a and 1b, respectively. List 2, four- and three-word categories are shown by 1c and 1d, respectively. The legend for all figures is the same as that in the upper left corner of 1a.)

three or four words of a category. On the other hand, the maximum recall for the NE categories occurred for two words of the three-word categories and three words of the four-word categories. Over both lists and all experiments, the percentage of the total component recall comprising occurrences of single items was as follows: E, four-word categories, 2.6%; E, three-word categories, 6.0%; NE, four-word categories, 7.6%; and NE, three-word categories, 16.7%. It is evident that if *S* can recall one word from a category there is a very high probability that he will recall at least two words from the category.

To determine the extent to which words of a category were recalled consecutively or not, an analysis of clustering was performed. Clustering is defined as the consecutive occurrence of words belonging to the same category (Bousfield, 1953). For the E and NE recall components, the percentage of the recall comprising occurrences of clusters of two, three, or four words was computed. Combining over all three experiments, the results were as follows: List 1—E categories, 89.57%; NE categories, 79.64%: List 2—E categories, 82.42%; NE categories, 67.20%. The overall percentage of clustering reflects the high degree to which *S*s grouped the recalled words into the appropriate categories of the lists. The mean difference between the number of clustered words of E and NE categories was 5.45 for List 1 and 5.60 for List 2. Both these mean differences were significant ($t = 11.85$, $p < .001$ and $t = 14.00$, $p < .001$ for Lists 1 and 2, respectively). The difference in clustering of words of E and NE categories between Lists 1 and 2 was not significant ($t = .24$).

The geometric means of the Thorndike-Lorge frequencies for the words of each category were correlated with an index of category usage. This index was derived by dividing the total recall of words of a given category by the maximum recall which could be obtained if all *S*s recalled all words of the category. In this manner, the indices of usage are independent of the fact that half the categories comprised three words and half four words. Separate rank-order correlations were computed for the E and NE components of each list for all three experiments. A test based on Fisher's z transformation (Johnson & Jackson, 1959) indicated that, for both lists, the correlations neither differed significantly from one another nor differed significantly from zero. The average correlation for List 1 was .100 and for List 2 was −.108. Thus, there is no evidence that frequencies of occurrence of the words of the categories influenced recall.

The next analysis concerned the relation between word recall per category and the index of interitem associative strength. For each category, the total word recall was divided by the number of *S*s who recalled at least one word of the category. This value was then converted into a percentage of the words comprising the category. The intercorrelations of this measure between the three experiments were extremely high—the Pearson r's ranged from .815 to .940. In view of this, the percentage maximum category recall measure for each category was averaged over each list and correlated with the interitem associative strength index. For List 1 the r was .784 while for List 2 it was .626. Both these r's are significant beyond the .01 level. It is apparent that the recall of words of a given category is highly related to the degree to which the words of the category elicit one another as word associates.

The mean interitem associative strength indices were as follows: List 1—E categories, 43.9; NE categories, 12.5: List 2—E categories, 22.4; NE categories, 5.6. Examination of the Table 2 recall means presented in Table 2 shows that the order of mean recall for each experiment is exactly reflected by the order of the mean interitem associative strength indices. Note, however, that though the interitem associative strength index reflects the within-category word recall, there is no consistent relation between this index and the number of categories represented by the recalled words.

Experiment IV.—In Exp. I, II, and III the use of categories as chunks to facilitate recall was inferred from the word recall data. In Exp. IV, *S*s were specifically asked to recall the names of the categories. For List 1, the mean number of categories listed by *S*s was 10.32 of which 5.16 were E categories and 5.16 were NE categories. For List 2, the mean number of categories was 10.30 of which 5.35 were E categories and 4.95 were NE categories. Neither the differences between the mean total number of categories for List 1 and List 2 nor the differences between mean numbers of E and NE categories for either list proved to be significant. Since the procedure used in Exp. III was most identical with that used for Exp. IV, the results of these two experiments were compared. The differences between the mean total number of categories for both lists as well as the differences between the mean number of E and NE categories for all comparisons were significant at better than the .05 level. Likewise, when the word recall for Exp. III and IV was compared, all differences were again significant at the .05 level or better.

The percentage category recall measure was computed for the word recall data. The intercorrelations of this measure between Exp. III and IV were .840 for List 1 and .954 for List 2. Also, the correlations between the percentage maximum category recall measure and the interitem associative strength index were .632 for List 1 and .548 for List 2. All these correlations are significant at better than the .01 level. Another comparison of interest between Exp. III and IV concerned the frequency with which the various categories were represented in recall. For List 1, the correlation was .619 while for List 2 it was .797. Again, these correlations are significant at beyond the .01 level.

Taken as a whole, the results of Exp. IV are in close agreement with the results of Exp. I, II, and III with the exception that the absolute magnitude of category and word recall is lower. Since *S*s in Exp. IV were not informed that they were to recall the categories (or for that matter, that they were to recall anything at all) until after the words had been exposed, it is reasonable to assume that no specific set was generated regarding recall of the material.

DISCUSSION

Whether measured directly or inferred from word recall, the experiments reported here consistently showed that only a limited number of categories (or categories of words) were handled in the free recall of the 20 category lists. Depending on the experiment, between 10 to 14 categories were recalled or represented in recall. This finding is compatible with results of free recall studies which utilized lists of unrelated words. Both Deese (1959a) and Murdock (1960) maintain that a portion of recall is based on the immediate memory span. Depending on various factors such as interitem associative strength, presentation time, etc., an additional number of items are recalled beyond the core provided by the memory span. Using Murdock's (1960, p. 231) formula, a recall of approximately 11 words is predicted from a list of 20 unrelated words presented at a 4-sec. rate. If the 20 categories of the present studies are equated with 20 unrelated words, the present findings are not at variance with this predicted value. Apparently, *S*s can handle the 20 categories of the list in essentially the same fashion as they handle a list of 20 unrelated words.

Several comments are required concerning the stable finding, over all the experiments, that despite the large difference in word recall between the E and NE categories, both category types, per se, are equally represented in recall. Since the interitem associative strength index is considerably higher for the E component of each list, it is evident that the efficacy of the index is restricted to word recall within a category and is not related to which categories will be recalled.

Miller (1956a) has made the distinction between information-rich and information-poor chunks. Information-rich chunks would be equivalent to the NE categories since the words of these categories represent only a small number of the alternative items belonging to the categories. Similarly, the E categories would be equivalent to information-poor chunks since the words of these categories can be chosen from only a small number of alternative items. On the basis of an information load interpretation, the poorer recall of the words of the NE categories would be expected to convey at least as much if not considerably more information than the E words. However, an information load analysis does not lead to a prediction of equal representation of E and NE categories of words.

The exact nature of a mechanism to (*a*) account for the equal use of E and NE word categories and (*b*) explicate the general nature of the recoding process is still a matter of conjecture. A promising approach to this problem may be based on the results of one of Deese's (1959b) recent papers. Deese has demonstrated that the extent to which an intrusion appears in free recall can be predicted from the average percentage occurrence with which the intrusion occurs as a free associate to the words of the list.

Extending this result to the present study, it may be assumed that each of the three or four words of a given category has a tendency to elicit not only other individual words as free associates but also a word or some recoded combination of words that represents something analogous to the postulated r_m-s_m mediator of Osgood, Suci, and Tannenbaum (1957) and Flavell (1961). If three or four words elicit the same or similar mediators, there should be a strong tendency for the mediator to remain active and aid in the organization and recall of the input information. That something of this nature occurred in this study may be inferred from a few instances where *S*s overtly intruded in recall a word which represented the category label for a specific group of words. For example, several *S*s intruded the words FLOWER and FISH in their recall protocols. FLOWER is an appropriate category label for the items DAISY, ROSE, and VIOLET while FISH is an appropriate label for the items PERCH, MINNOW, SALMON, and SARDINE.

REFERENCES

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. Cultural norms for verbal items in 43 categories. Technical Report No. 22, 1957, University of Connecticut, Contract Nonr-631(00), Office of Naval Research.

DEESE, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312. (a)

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22. (b)

FLAVELL, J. H. Meaning and meaning similarity: I. A theoretical reassessment. *J. gen. Psychol.*, 1961, **64**, 307–319.

JOHNSON, P. O., & JACKSON, R. W. B. *Modern statistical methods.* Chicago: Rand McNally, 1959.

MILLER, G. A. Information and memory. *Scient. American*, 1956, **195**, 42–46. (a)

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97. (b)

MURDOCK, B. B. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois Press, 1957.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received May 23, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 377–384

EFFECTS OF RESPONSE MEANINGFULNESS (*m*) ON TRANSFER OF TRAINING UNDER TWO DIFFERENT PARADIGMS[1]

JOHN JUNG[2]

Northwestern University

The effects of 2 levels of response meaningfulness (m) on transfer under 2 paradigms, A-B, C-B and A-B, A-C were studied using paired-associate lists of 6 pairs of items consisting of adjectives as stimuli and trigrams as responses. Since the same responses occur on both lists of the A-B, C-B paradigm, positive transfer or facilitation of 2nd list learning may result from response learning on the 1st list. Such transfer should be greater with low m responses which require more response learning than high m responses. The results supported the prediction of greater positive transfer with low m. Although negative transfer was predicted for the A-B, A-C paradigm, it was predicted to be independent of Response m. However, high m produced greater negative transfer.

The learning of paired-associate lists has been logically analyzed into two major components, a response-learning stage and an associative stage (Hovland & Kurtz, 1952; Underwood, Runquist, & Schulz, 1959). During the first stage the responses per se are learned so that they become readily available units. Secondly, these responses must be associated to the appropriate stimuli with which they have been paired.

Ellis and Burnstein (1960) have advocated extending the above analysis of the learning of a single paired-associate list to the learning of two successive paired-associate lists, i.e., the transfer of training paradigm. They analyzed the A-B, C-B paradigm (stimuli different, responses identical) accordingly. Since the responses on the two lists are identical, response learning which occurs on the first list should be readily transferred to the learning of the second list. Learning of the second list can begin with the second or associative stage since the responses necessary were already learned during training on the first list. Thus, positive transfer is possible via the response-learning stage in this paradigm.

Adding to the Ellis-Burnstein analysis, it can be seen that negative transfer may occur under this paradigm via the associative stage. This transfer task requires the learning of two different associations with the same responses and it is not inconceivable that interference may be encountered in such a situation. In fact, Twedt and Underwood (1959) obtained a slight amount of negative transfer using this paradigm. They attributed this result to possible interference of backward associations from the first list to the forward associations of the second list. Thus, whereas the response-learning stage produces an element of positive transfer under this paradigm, the associative stage leads to negative transfer effects. Transfer effects obtained under this

---

[1] This study is based on a PhD dissertation submitted to the Graduate School of Northwestern University. The author is indebted to Benton J. Underwood for his guidance and encouragement during the course of the investigation.

[2] Now at Long Beach State College, Long Beach, California.

TABLE 1

MEAN NUMBER OF ANTICIPATION TRIALS TO CRITERION ON LIST 1

Paradigm	Response m	
	High	Low
A-B, C-B	9.50	24.90
A-B, A-C	10.50	23.45
A-B, C-D	10.85	23.40

were no significant differences in first-list learning as a function of paradigms under either high m, F (2, 57) = .37, or under low m, F (2, 57) = 1.15.

List 2 learning, Trials 1–2.—Figure 1 presents the mean number of correct anticipations for each condition on the first two anticipation trials of List 2. A greater number of correct anticipations was found with high than with low m. The F for Response m of 61.80 with 1/114 *df* was significant at the .001 level. The relative performance of groups under the different paradigms was identical at both levels of Response m, C-B being highest, followed by C-D, and finally A-C. The differences between the paradigms was significant, F (2, 114) = 16.67, $p < .001$. The Response $m \times$ Paradigm interaction was also significant, F (2, 114) = 4.17, $p < .05$, reflecting the fact that the magnitude of the effect of Response m is not the same for all paradigms.

In order to measure specific transfer, it is necessary to partial out any nonspecific transfer effects which are derived from warm up, etc., in proceeding from the first list to the second. In the present study, this is possible by comparing the C-D control (which provides a measure of nonspecific transfer) with each of the other two paradigms (A-C and C-D). The latter paradigms may produce some transfer attributable to specific stimulus-response relationships in addition to nonspecific transfer.

Inspection of Fig. 1 indicates that

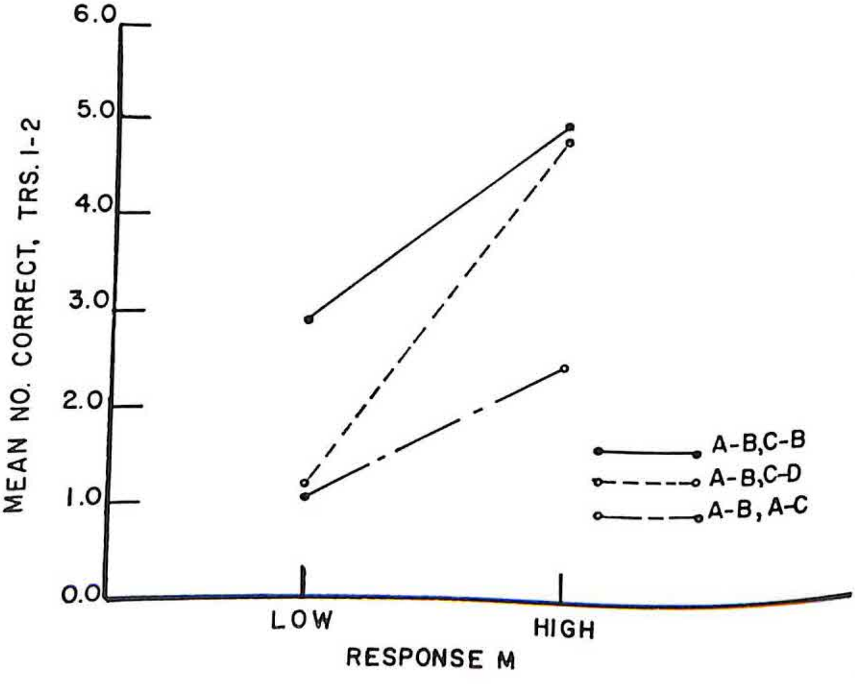


FIG. 1. Mean number of correct anticipations as a function of Response m and paradigms on Trials 1–2.

the superiority of the performance of the C-B condition to that of the C-D condition is greater with low than with high *m*. This finding shows that positive transfer under this paradigm is greater with low than with high *m*, as predicted earlier.

To determine if these differences in net transfer as a function of Response *m* were significant statistically, difference scores were obtained between the number of correct anticipations for each *S* in the C-B paradigm and the mean number of correct anticipations for the C-D condition at each level of Response *m*. First, the mean number of correct anticipations was obtained for the C-D condition at each level of Response *m*. Then, for each *m* level, the C-D mean score was subtracted from each C-B score at that level of *m*, providing two distributions of difference scores. A t of 2.63 with 19 *df* was obtained between the difference scores obtained for high and low *m* ($p < .01$).

A similar analysis was made of transfer under the A-C paradigm. Figure 1 indicates that the inferiority of the A-C condition to the C-D condition was greater with high than with low *m*. In other words, net negative transfer was obtained under the A-C paradigm and it was greater with high than with low *m*, contradicting the hypothesis that Response *m* would not differentially affect transfer under this paradigm. Statistical analysis was performed in the same manner described above for the the C-B paradigm. The difference scores obtained for each level of Response *m* differed very significantly, t (19) = 5.38, $p < .001$.

List 2 learning, Trials 1–10.—The mean number of correct anticipations on all 10 trials of List 2 for each condition is presented in Fig. 2. For all paradigms, better performance occurred under high than under low *m*, F (1, 114) = 79.95, $p < .001$. Performance also differed significantly as a function of paradigms, F (2, 114) = 18.61, $p < .001$. For low *m*, the highest performance came under C-B, followed by C-D, and finally by A-C; however, under high *m*, performance was highest under C-D, then C-B, and finally, A-C. This difference in the rank ordering of the paradigms as a function of Response *m* is reflected by a significant Response *m* × Paradigm interaction, F (2, 114) = 10.40, $p < .001$.

As in the case of the results for Trials 1–2, to obtain measures of net transfer it is necessary to compare the C-D condition with each of the other two paradigms in performance on the transfer list. Inspection of Fig. 2 indicates that the superiority of the C-B paradigm to the C-D paradigm on performance on Trials 1–10 is, as it was on Trials 1–2, greater under low than with high *m*. Actually, with high *m*, the performance of the C-B condition is inferior to that of the C-D paradigm.

Difference measures were obtained by comparing the mean number of correct anticipations for the C-D condition with the number of correct anticipations for each *S* in the C-D paradigm at each level of Response *m*. The difference scores for the high and low *m* groups were significantly different, t (19) = 3.82, $p < .01$. These findings are in agreement with similar findings from Trials 1–2 and support the prediction based on the response-integration hypothesis that transfer will be greater with low than with high *m* in this paradigm.

Figure 2 also indicates that the inferiority of the A-C paradigm to the C-D paradigm in performance on Trials 1–10 is greater under high than under low *m*. This finding agrees with a similar relationship found between Response *m* and net transfer

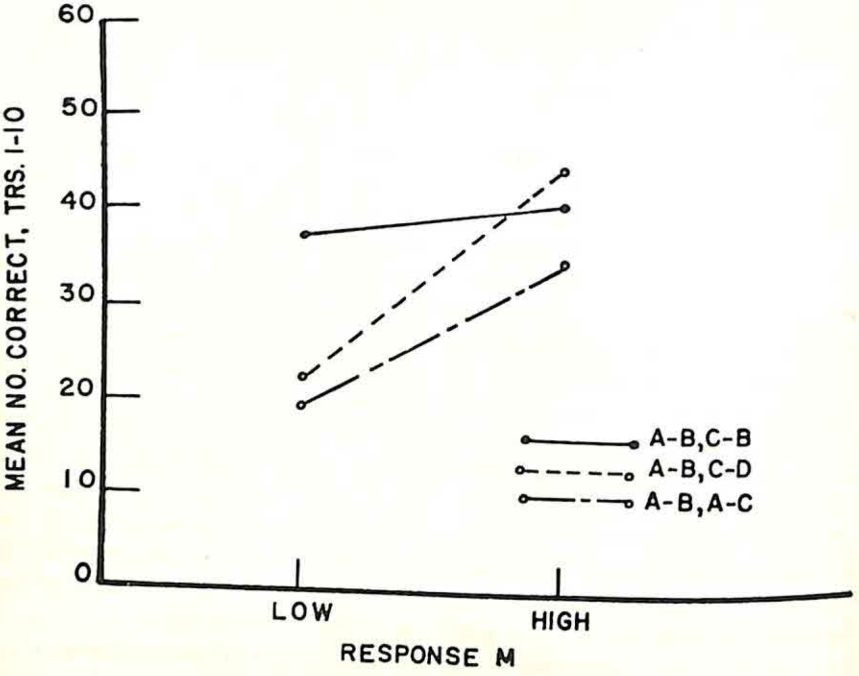


FIG. 2. Mean number of correct anticipations as a function of Response *m* and paradigms on Trials 1–10.

on Trials 1–2 although it is less striking on Trials 1–10. Difference scores were obtained to provide measures of net transfer at each level of Response *m* in the same manner used for the C-B paradigm described above. The *t* of 2.25 ($df = 19$, $p < .05$) obtained showed that net negative transfer was significantly greater with high than with low *m*. This result, like the relationship found on Trials 1–2, is incompatible with the earlier prediction that transfer under this paradigm is independent of Response *m*.

Error analysis.—In learning a paired-associate list it is reasonable to expect that as the course of training proceeds erroneous responses will increase up to some point and then decrease as learning becomes complete. The portion of training during which errors gradually increase to some peak corresponds roughly to the response-learning stage; the period of training during which errors decrease parallels the associative-learning stage.

One may analyze the transfer of training situation in terms of how the learning of a prior list affects the error-making tendencies on the learning of the second list. In the C-B paradigm, since the responses which appear on the second list have been integrated on the first list, little or no response learning is necessary at the beginning of learning of the second list. On the other hand, for the A-C and C-D paradigms, a completely different set of responses is encountered on the second list. Response learning as well as associative learning is necessary on the second list for these two paradigms.

The preceding analysis implies that the course of error committment on the transfer list for the C-B paradigm will be different from that under the A-C and C-D paradigms. To be explicit, error tendencies should be maximal at the very outset of List 2 learning for the C-B paradigm and minimal for the A-C and C-D paradigms. As List 2 training proceeds, errors will steadily decrease until they are eliminated for the C-B paradigm; on the other hand, errors will first increase to some point before decreasing and finally become eliminated for the A-C and C-D paradigms. Not only should the shapes of the error curves be different for the C-B as opposed to the A-C and C-D paradigms but these differences should be affected by the level of Response *m*. The higher the response *m*, the less should be the difference in the shapes of the error curves between the C-B and the A-C and C-D paradigms because response learning is relatively fast. In other words, the advantage of having responses on List 2 which have already been integrated on List 1 for the C-B paradigm is minimized with higher degrees of Response *m*. On the other hand, with low *m*, the differential effects of response integration should be greater and lead to the expected differences in the shapes of the error curves for the C-B as opposed to the A-C and C-D paradigms.

It is not possible to meaningfully interpret the absolute or even the relative frequencies of errors under the various conditions since the degree of learning was not equated. This problem is avoided by restricting the interpretation of errors to the shapes of the error curves. Figure 3 presents the course of intrusion errors on List 2 learning for both levels of Response *m* and for each type of paradigm (C-B compared to A-C and C-D). An intrusion is an error which is correct as a response but given to an inappropriate stimulus. As predicted, under low *m* the commitment of intrusion errors decreased over training for the C-B paradigm whereas it increased over training for the A-C and C-D paradigms. This finding is consistent with the finding of maximal positive transfer under low *m*. But as also shown in Fig. 3, the differences in the shapes of the error curves for the two types of paradigms are less striking under high *m* than they were under low *m*. Errors decreased over training for the C-B paradigm as they did under low *m*; for the A-C and C-D paradigms, errors increased at first but reached a peak quite early and began to decrease over the remainder of List 2 training whereas under low *m* errors were still increasing on the last block of List 2 trials. These findings are

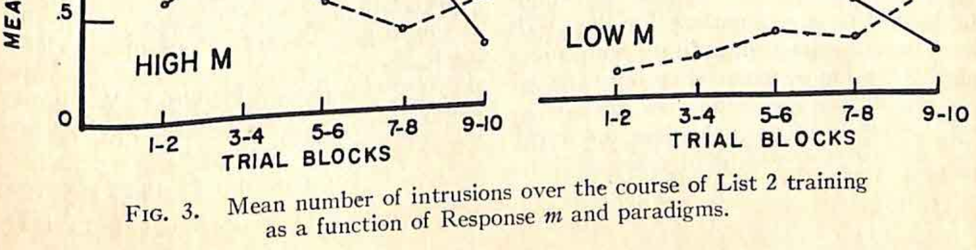


FIG. 3. Mean number of intrusions over the course of List 2 training as a function of Response *m* and paradigms.

TABLE 1

AGE, HEIGHT, AND INTERPUPILLARY DISTANCE

Group		Age	Height (In.)	Interpupillary Distance (Mm.)
5½	Range Mdn.	4-9 to 6-5 5-6	42½ to 50 47	
7	Range Mdn.	6-8 to 7-10 7-2	45½ to 55 50	49 to 55 54
10	Range Mdn.	9-6 to 10-6 9-11	52½ to 61 56	48 to 56 54
12	Range Mdn.	11-5 to 12-3 11-9	55½ to 64 57½	50 to 60 55
Adult	Range Mdn.	18 to 30 23	63¼ to 71½ 68½	

age and the oldest was 30 yr. old. The median and range for each group are given in Table 1 which also gives similar data as to the *S*s' heights and interpupillary distances. The youngest age group, it should be noted, covered a range of almost 2 yr. from 4-9 to 6-5. The other children's groups have a narrower age range spanning approximately 1 yr. There were 12 *S*s in each group, 6 males and 6 females. The *S*s included children attending a preschool nursery; children attending a suburban elementary school; and an adult group comprised of graduate students, medical students, technical assistants, and secretaries.[2] The *S*s' visual acuity was at least 20/20 corrected.

Judgments of distance were made in a level, grassy field next to the school where the *S*s were enrolled or where they were employed. The open terrain to the front of the *S*s extended for at least 65 ft. without any clear marker such as a bush or path, but this varied for the different groups. Thus, for the youngest group and for the adults the field extended 65 ft., for the 7-yr.-olds it extended about 100 ft., for the 10-yr.-olds and for the 12-yr.-olds there was about a 75-ft. expanse.

The *S*s stood at one end of the field and directed the *E* to mark off successive 1-ft. intervals using as a standard a foot-rule placed on the ground at the foot of and immediately to the front of the *S*. The instructions were:

> You see the ruler on the ground, I am going to move the pointer along the ruler and I want you to tell me to stop when I have moved the pointer the length (to the end) of the ruler. Now I'll put this marker (pencil) at this point. Now I'll start from the marker and again you tell me to stop when I've moved the pointer as far (as long) from the marker as the ruler is long.

The marker was a yellow pencil and the pointer which was moved by the *E* was a long metal rod. After each judgment the *E* measured the distance from the marker to the pointer and moved the marker to the spot where the pointer had been stopped. The next judgment was begun then. Two *E*s were present so that each *S* could be watched as they made their judgments. If an *S*'s attention wandered, as occurred at times with the younger children, the child's attention was recalled to the task. If there was doubt that the child truly was attending, he was asked if he wanted the *E* to stop moving the pointer or if he wanted it moved forward or back. On a few occasions a particular judgment was begun again, but this was never done after the *S* had completed a judgment.

Each *S* gave two series of judgments: one at normal height, the other at an adjusted height. The adjusted height judgments for the four groups of children were obtained while each child was standing on a platform such that his (her) overall height was 5 ft. 6½ in. (plus or minus ¼ in.). The adjusted height judgments for the fifth group, the adults, were obtained while the *S*s were

---

[2] The children in the youngest group were obtained through the cooperation of Elizabeth M. Staub, Director, Nursery School, Department of Psychiatry, University of Rochester. The three older groups of children were obtained through the cooperation of the Brighton, New York schools, District No. 1 and Fred B. Painter, Superintendent; Elsie Robinson, Principal; and Francis E. Davies, Principal.

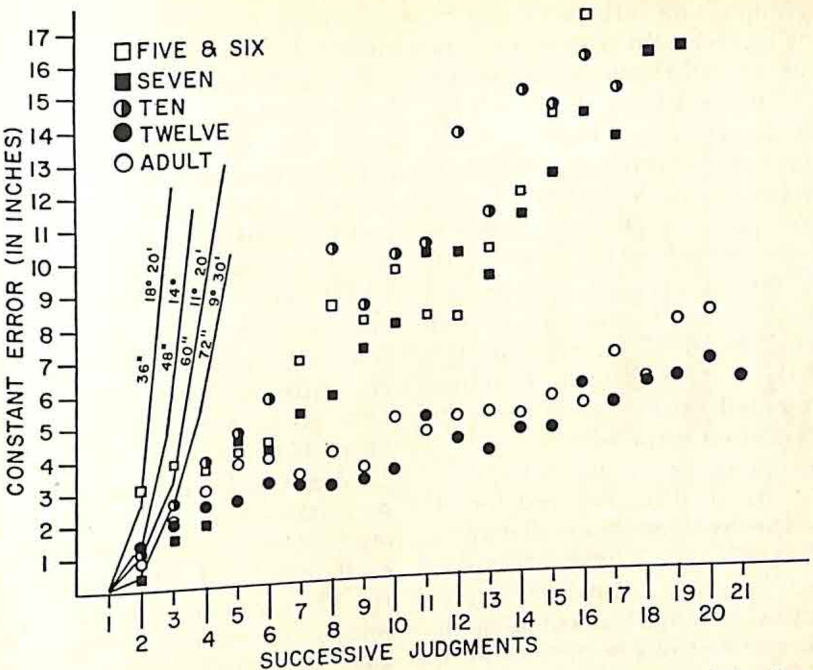


FIG. 1. Constant error for successive distance judgments. (The line curves at the left represent visual angle functions for a hypothetical *S* whose eyes are 36, 48, 60, and 72 in. above the ground.)

kneeling. Kneeling height ranged from $33\frac{3}{4}$ in. to 53 in. with a median of $47\frac{1}{4}$ in. Thus, the adjusted height of the adult group was in the opposite direction from that for the four groups of children. Also, the extent of the adjustment was greater for the shorter and thus mainly for the younger children. Within each age and sex grouping half the *S*s made normal height judgments first and half the adjusted height judgments first.

Initially, the hope had been to obtain 19 successive 1-ft. estimates for each *S* at each height, that is, estimates starting with the second foot-length distal from the *S* and extending to the estimated foot-length at a perceived distance approaching 20 ft. from the *S*. This was not possible in each instance. Some of the *S*s' estimates were so great that the total linear expanse of the field had been included before they were at the nineteenth estimate. The smallest number of judgments obtained from an *S* for a given height was 8. For the youngest group, though all understood the instructions (repetition of instructions was necessary at times), not all were able to tolerate the time and effort required. Fatigue and distractibility in 5 *S*s necessitated our dropping them from the sample and replacing them with other children. An examination of the data available on these five children, however, indicates that their performance was not different from that of the children who were able to attend sufficiently.

RESULTS

Each judgment by an *S* provides an estimate of the extent to which he over- or underestimates the 1-ft. interval. If a given interval is overestimated, the distance marked off for a specific judgment should be less than 1 ft.; if the distal interval is underestimated, greater than 1 ft. should be judged equal. As the judgment is made at greater distances from the *S*, a larger extent is required to provide an equal angle at the retina. Thus underestimation is equivalent to underconstancy, that is, a judgment falling between the requirements of constancy and of visual angle.

Figure 1 shows the average underestimation (constant error) for each successive normal height judgment by age group. The vertical axis gives the error in inches by which a given judgment exceeded the standard of 12 in. Only data from those *S*s who made judgments at their normal height on the first series are included. The curves at the left of the graph represent visual angle functions for those instances when an *S*'s eyes are 36, 48, 60, and 72 in. above the ground. The angle indicated is that subtended by a 12-in. ruler at the foot of the *S*. Constancy, of course, is represented by the baseline of the graph, constant error = 0.

Three facts are immediately apparent. At all distances and for all groups the average error underestimates the distance so that the group average indicates something more than 12 in. is judged as equaling the ruler at the foot of the *S*. Some *S*s, particularly at nearer distances, did give judgments of less than 12 in. but the group trend is clear. It is also evident that the farther the distance from the *S*, the greater the underestimation of the interval. Third, on gross inspection the slope of increasing error with successively more distant judgments seems similar for the three younger groups and steeper than for the two older groups.

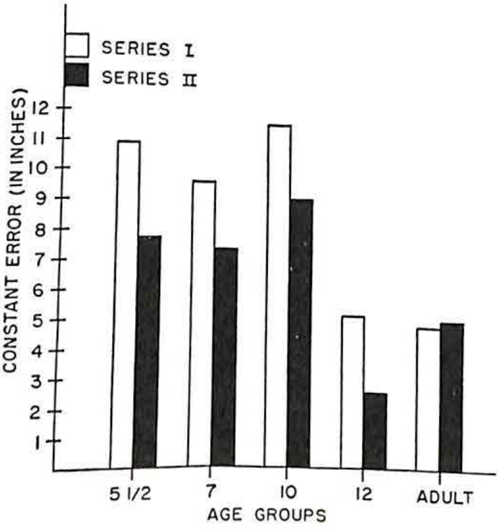


FIG. 2. Constant error for each age group on Series I and Series II.

Using the average underestimate for all judgments of a given series as a measure descriptive of an *S*'s performance, an analysis of variance for age, sex, height (normal or adjusted), and order (first or second series) indicates that the height at which the judgment is made does not influence the result.[3] Sex, likewise, is not a pertinent variable. Age significantly affects this measure ($F = 2.80$, $p < .05$) as does order ($F = 37.63$, $p < .01$). There also is an Age×Order interaction ($F=3.18$, $p < .05$). The three younger groups are not significantly different from each other and all three differ from the 12-yr.-old and adult groups. Age means for each order are shown in Fig. 2. There is an increase in veridicality of judgments for the second series for all groups except for the adults; thus, the significant interaction as well as the main effect due to order.

The increasing separation among the younger and older *S*s' performances with more distal judgments evident in Fig. 1 and the fact that age differences in size constancy (which relates size to distance) apparently are evident at farther distances (Zeigler & Leibowitz, 1957) suggested the desirability of repeating the analysis using average underestimation scores for the second half (farther distances) of each series of judgments. There is no change in our results thus confirming that the age differences already mentioned are the only ones present.

The height at which the judgments were made does not influence the *S*'s response. The lack of effect of height for each group is shown in Fig. 3. The apparent shift for the adult group is not significant. It should be reiterated that the adjusted height for the adults was obtained by having them kneel, thus reducing their height while the adjustment for the other groups was to raise them above their normal height. Height as a possible influence on distance judgment can be evaluated also within each age group by correlating the individual's height with his performance. The correlation of underestimation and *S*'s height, with age held constant, is −.075.

---

[3] I am grateful to Russel F. Green for his help with the analysis of variance.

Interpupillary distance was measured in three of our *S* groups: 7-, 10-, and 12-yr.-olds. A correlation of interpupillary distance with extent of underestimation, holding age constant, is −.206 and is not significant statistically.

The variance within each age group decreases as the age of the group increases. The possibility that the greater variability in the younger groups was a function of lack of attention or of not understanding instructions was checked by correlating the error scores for the first series with those for the second in each group. If the younger children were inattentive or found the task too difficult, one would expect little correlation between the two series of judgments or, at least, less correlation than that obtained by the older children and adults. The respective product-moment correlations starting with the youngest group are: .93, .96, .93, .87, .57.

DISCUSSION

Spatial parameters such as angle of regard appear to play a complex role in distance perception. As distance from the *S* increases, there is greater underestimation of the distal extent at all age levels investigated in the present study. At farther distances, an interval larger than 1 ft. is required for the *S* to give a judgment of equality to a foot-rule immediately in front of him; and as distance from the *S* increases a greater interval is required. While this is true at all ages, the strength of this effect is reduced significantly between the ages of 10 and 12. That this reduction is not primarily a function of the change in height with age is evident.

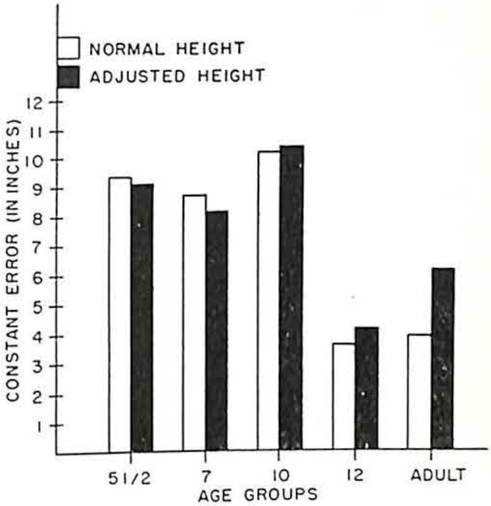


FIG. 3. Constant error for each age group at normal height and adjusted height.

A comparison of our results with Gilinsky's (1960) shows reasonable agreement for normal height binocular estimates. In her study she demonstrated clear differences between judgments obtained under monocular as compared with binocular regard, the latter having greater veridicality. In the current study, we did not experimentally manipulate this variable so that no direct evidence on this point can be offered. While inter-*S* differences in interpupillary distance had no influence, this is both quantitatively and qualitatively different from reduction to monocular viewing. Our results are consistent with Mead's (1943) finding of a lack of significant correlation between interpupillary distance and stereopsis score in a test of stereoscopic depth perception and with Clark and Warren's (1935) conclusion that interpupillary distance and depth perception as measured by the Howard-Dolman apparatus are not related.

A geometric explanation for the developmental trend then is not possible. The simple invocation of experience, likewise, does not appear to be sufficient. If one assumes increasing familiarity and experience with space with increasing age, one would expect a more gradual increase in distance constancy than is demonstrated. While it would be attractive to explain the shift between the ages of 10 and 12 as a function of emerging spatial concepts influencing earlier perceptual response, (Piaget & Lambercier, 1951, 1956), the problems for constructing an adequate developmental hypothesis are compounded by the great variability in the younger groups. The most accurate of the 5½-yr.-old *S*s performs better on the first series than all of the older *S*s except for one 12-yr.-old. There is greater homogeneity among the members of a particular age group, the older the group. It is unlikely that this variability is due to difficulty in the younger children understanding the instructions as even the child producing the greatest errors exhibited a marked consistency in his judgments. It will be necessary to examine such individual differences before a reasonable conclusion as to the processes involved can be entertained.

The data of this study suggest that for this procedure, repeated exposure to the task leads to increased perceptual differentiation (Gibson & Gibson, 1955) in the children but not in the adults. The second series of judgments, in the absence of information as to the accuracy of the initial series of judgments, improves for all groups except the adults. This is not a result of the greater initial accuracy of the adults as the 12-yr.-olds do equally well the first time yet clearly do improve. Conceivably the different finding for the adult group is a function of their dissimilar experimental manipulation, though there is no immediately obvious reason why this would lead to such a result.

We are still some distance from an adequate analysis of the demands which a particular task places on different *S*s. We cannot say that the results reported here would recur if distance judgments were examined by the method of fractionation (Purdy & Gibson, 1955) or if the focus were on interspaces (Costa, 1949). A comparison of developmental patterns among these and other methods, obtained from *S*s drawn from similar populations, would help define the relevant parameters.

REFERENCES

CLARK, B., & WARREN, N. Depth perception and interpupillary distance as factors in proficiency in ball games. *Amer. J. Psychol.*, 1935, **47,** 485–487.

COHEN, W., HERSHKOWITZ, A., & CHODACK, M. Size judgments at different distances as a function of age level. *Child Develpm.*, 1958, **29,** 473–479.

COSTA, A. M. Constanza a percettiva di intervalli vuoti. *Arch. Psicol. Neurol. Psichiat.*, 1949, **10,** 377–388.

EPSTEIN, W., PARK, J., & CASEY, A. The current status of the size-distance hypotheses. *Psychol. Bull.*, 1961, **58,** 491–514.

GIBSON, J. J., & GIBSON, E. J. Perceptual learning: Differentiation or enrichment? *Psychol. Rev.*, 1955, **62,** 32–41.

GILINSKY, A. S. Perceived size and distance in visual space. *Psychol. Rev.*, 1951, **58,** 460–482.

GILINSKY, A. S. The effect of growth on the perception of visual space. Paper read at Eastern Psychological Association, New York, April 1960.

MEAD, L. C. The influence of size of test stimuli, interpupillary distance, and age on stereoscopic depth perception. *J. exp. Psychol.*, 1943, **33,** 148–158.

PIAGET, J., & LAMBERCIER, M. Recherches sur le développement des perceptions: XII. La comparaison des grandeurs projectives chez l'enfant et chez l'adulte. *Arch. Psychol., Geneva*, 1951, **33,** 81–130.

PIAGET, J., & LAMBERCIER, M. Recherches sur le développement des perceptions: XXIX. Grandeurs projectives et grandeurs réeles avec étalon éloigné. *Arch. Psychol., Geneva*, 1956, **35,** 257–280.

PURDY, J., & GIBSON, E. J. Distance judgment by the method of fractionation. *J. exp. Psychol.*, 1955, **50,** 374–380.

ZEIGLER, H. P., & LEIBOWITZ, H. Apparent visual size as a function of distance for children and adults. *Amer. J. Psychol.*, 1957, **70,** 106–109.

(Received May 31, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 391–397

PREDICTION OF BEHAVIOR UNDER CONDITIONS OF UNCERTAINTY

JULIAN O. MORRISSETTE AND WILLIAM H. PEARSON

Behavioral Sciences Laboratory, Wright-Patterson Air Force Base, Ohio

Erratic lines approaching a target space were presented to Ss and described as paths of aircraft approaching a complex of (9) targets. Ss predicted the 3 targets each path was most likely to intersect. S performance (predictions) was evaluated against the probability of each target being intersected by a path, where this probability was computed mathematically. By sampling from an information source, whose entropy, H, was held constant, 10 basic paths of varying sample entropy, $\hat{H}$, were constructed. Each of the 10 paths was truncated at 3 different lengths (in effect, 3 different sample sizes), yielding 30 paths, each with a unique $\hat{H}$. 2 hypotheses were proposed: (a) S performance would be superior on samples of low $\hat{H}$ as compared with samples of high $\hat{H}$, and (b) S performance would improve as sample size increased. The analysis of the data produced statistically significant differences consistent with the hypotheses.

Following the work of Shannon and Weaver (1949), psychologists have shown a persistent interest in information theory. In general, information theory has been used in the service of two problems: (*a*) determination of the capabilities of the human operator as an information transmitter (e.g., Eriksen & Hake, 1954; Hake & Garner, 1951; Klemmer & Frick, 1953), and (*b*) describing the characteristics of tasks, i.e., complexity, difficulty, redundancy, etc. (e.g., Attneave, 1955; Berlyne, 1957; Fitts, Weinstein, Rappaport, Anderson, & Leonard, 1956).

The experiment to be reported here is concerned with the latter problem. In the present work an aspect of information theory is employed that has not, apparently, been used before by psychologists. Therefore, a brief development of the mathematics involved seems called for.

Information theory defines the entropy (degree of randomness of events) of stochastic processes (a sequence of events governed by a set of probabilities). A special type stochastic process is a Markov chain (the interest of this paper), where the probability of the event e_n is dependent upon the event e_{n-1} and only e_{n-1}. The statistical structure of such an event scheme can be represented by a set of probabilities, $p_{i(j)}$, the probability that the event i is followed by the event j.

The formula for computing the entropy of Markov chains is given by[1]

$$H = -\sum_{i,j} P_i p_{i(j)} \log p_{i(j)} \qquad [1]$$

H may also be interpreted as representing the average amount of information associated with each event in a sequence. Dropping the weighting factor, $P_i p_{i(j)}$, from Equation 1 and considering each event individually we have:

$$h_i = -\log p_{i(j)} \qquad [2]$$

here h_i is the amount of information associated with each event (Attneave, 1959, p. 6). Summing over a set of observed h_i's, we have:

---

[1] P_i in Equation 1 refers to "stationary probabilities." These will be discussed more fully later in the text.

$$h_T = \sum n_i h_i \qquad [3]$$

the total amount of information in a sequence of events. When h_T is divided by N, the number of events in a sequence, we have:

$$\hat{H} = h_T/N \qquad [4]$$

which is the entropy of the sample and a sampling statistic.

As a sampling statistic, two points may be noted about $\hat{H}$: (*a*) repeated samples of size N from an information source will yield values of $\hat{H}$ which will be normally distributed about H (Miller & Madow, 1954, p. 2),[2] and (*b*) in general, the larger the sample size N, the better will $\hat{H}$ be an estimate of H (Mood, 1950, p. 149). It should be noted that in computing $\hat{H}$ that the probabilities, $p_{i(j)}$, are the known probabilities of events occurring in the information source, not estimates ($\hat{p}_{i(j)}$) of them.

One more point regarding Equation 1 must be clarified: namely, the meaning of P_i. In discussing stochastic processes, the convention is to call the events emitted by an information source *states*. Thus, one may talk about the *states of the system* and the probabilities of the source being in any given state at time t. In dependent stochastic processes these probabilities P_i, one associated with each state and constant for all trials for a given information source, are called the *stationary probabilities*. The process of computing stationary probabilities is too involved to present here, therefore, the reader is referred to Kemeny, Snell, and Thompson (1957, pp. 217 ff.).

We now consider the implications of these developments for research purposes. We note first that previous psychological studies involving information theory may be described as follows: (*a*) h_T/N always equaled H since the frequency with which events were presented to *S*s was controlled so as to correspond with the experimentally defined probabilities of each event occurring; (*b*) task properties were manipulated by changing H; and (*c*) the amount of information presented, h_T (the number of presentations of stimuli), has been left uncontrolled. From the point of view of the present development, sample entropy, $\hat{H}$, and sample size, N, are significant variables. Thus, to explore the usefulness of describing tasks in the language of sample entropy and sample size, an experiment was designed in which these variables were manipulated and human performance assessed.

METHOD

The task.—Ten paths were reproduced on 8 × 10 in. paper, with nine potential "targets" enclosed in a square near the edge of the paper. The 10 paths were samples drawn from an information source, and these were varied by three different lengths, representing variations in sample size, thus producing 30 paths each with a unique $\hat{H}$. In order to determine if path direction effects existed, the direction of the paths was varied by presenting each path in one of four rotated positions, so that the target area appeared either at the upper right or left or lower right or left. In all, then, there were 120 paths in the stimulus material: 10 basic paths × 3 path lengths × 4 positions of target area.

Figure 1 is a sample of the stimulus ma-

---

[2] The only situation where this is not true is in the case where all events are equally likely. Here $h_T/N = \hat{H} = H$, irrespective of sample size, N, and the distribution of events in the sample. This is true since in the equally probable case $H = -\log p_i$ and $h_i = \log p_i$ (Attneave, 1959, pp. 5–6), and $h_T/N = (h_1n_1 + h_2n_2 + \cdots + h_kn_k)/N$, where $n_1 + n_2 + \cdots n_k = N$ and $h_i = h_j$: therefore, $(h_1n_1 + h_2n_2 + \cdots + h_kn_k)/N = h_i(n_1 + n_2 + \cdots + n_k)/N = h_i$. In an information source where all events are equally likely, a point distribution of $\hat{H}$ is obtained.

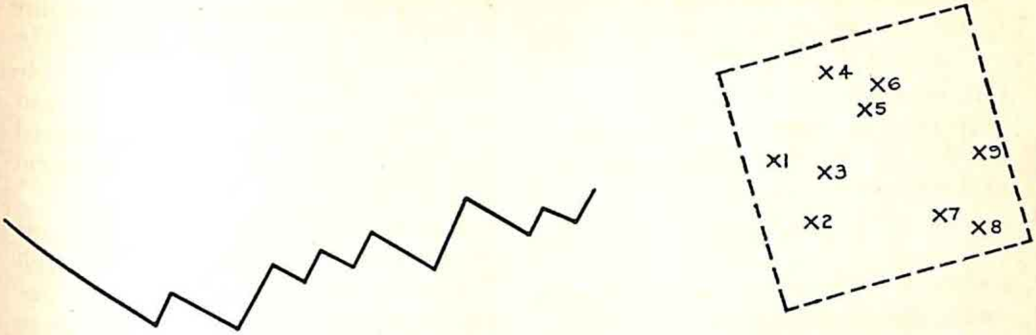


FIG. 1. Sample of path presented to *S*s.

terials presented to the *S*s. So that an experimental session could be completed in 1 hr., thereby minimizing the effects of boredom and fatigue, the reproduced paths were bound in two books, each containing 60 paths. Assignment of paths to books, as well as sequence within books was random.

Procedure.—Two experimental sessions were run, with 30 *S*s in one group receiving one set of 60 paths, and 27 *S*s in the second group receiving the other set of 60 paths. The paths were described to *S*s as representing a crude simulation of a set of pictures which a radar monitor might receive of an aircraft flying an evasive pattern toward a set of targets all of equal strategic value. The *S*s were asked to consider each path and from the information contained in the path predict the three most likely targets. The *S*s were given 30 sec. to make their decision and record their predictions in their books.

Subjects.—The *S*s used in the experiment were male students from the University of Dayton, Dayton, Ohio. All *S*s were volunteers and they were paid for their participation.

Technical discussion of the task.—To obtain the 10 basic paths an "event field" was mapped out on 10 × 10 cm. graph paper. On the ordinate 21 states, 1–20, were indicated, representing the possible states of the system, while on the abscissa 41 steps, 0–40, were indicated, representing a sequence of states of the system.

To define the set of transition *possibilities* from one state to another for each of the 21 states, the following rule was used: If State i occurs at Step n, then at Step $n+1$, only the state $i+1$ or $i-1$ will occur. Remaining at State i was excluded. To the set of transition *possibilities* permitted by this rule a set of transition *probabilities* was assigned according to the following rule: the probability $p_{i(j)}$, of transition from State i to State j, where $j = i + 1$ is $1 - i/20$, while the transition probability, $p_{i(j)}$, where $j = i - 1$ is $i/20$. The transition possibilities and their associated transition probabilities are shown in Table 1, along with h_i for each $p_{i(j)}$. From the transition probabilities the stationary probabilities, P_i, the probabilities of the system being in any given state at any step, n, were computed, and these are shown in Table 2.

From the rules of transition and the transition probabilities, paths were drawn. The initial state (0–20) for a path was selected

TABLE 1

TRANSITION POSSIBILITIES, TRANSITION PROBABILITIES, $p_{i(j)}$, AND $h_i = -\text{LOG } p_{i(j)}$

Transition Possibilities		$p_{i(j)}$	h_i	Transition Possibilities	
From State	To State			From State	To State
0	1	1.00	0.000	20	19
1	0	.05	4.322	19	20
1	2	.95	0.074	19	18
2	1	.10	3.322	18	19
2	3	.90	0.152	18	17
3	2	.15	2.737	17	18
3	4	.85	0.234	17	16
4	3	.20	2.322	16	17
4	5	.80	0.322	16	15
5	4	.25	2.000	15	16
5	6	.75	0.415	15	14
6	5	.30	1.737	14	15
6	7	.70	0.515	14	13
7	6	.35	1.515	13	14
7	8	.65	0.622	13	12
8	7	.40	1.322	12	13
8	9	.60	0.737	12	11
9	8	.45	1.152	11	12
9	10	.55	0.836	11	10
10	9	.50	1.000	10	11

TABLE 2

Stationary Probabilities, P_i, of the 21 States

States	P_i	States
0	0.000	20
1	0.000	19
2	0.000	18
3	0.001	17
4	0.005	16
5	0.015	15
6	0.037	14
7	0.074	13
8	0.120	12
9	0.160	11
10	0.176	

from a table of random numbers, giving each state an equal chance of being the point of origin.[3] Given the initial state of a path, Step 0, State *i*, the determination of the state of the path at Step 0 + 1, and all succeeding steps, were based on a chance selection from a table of random numbers.

This procedure was repeated 10 times in order to draw a sample of 10 paths. In order to manipulate sample size each of the 10 paths was "cut" off at three different lengths, 27, 24, and 21 steps, yielding 30 paths. Table 3 contains the $\hat{H}$ values of the 10 paths for each sample size, and the entropy of the information source from which the paths were drawn.[4]

Each of nine targets was assigned to a different state within the target area (6, 7, . . . , 14) and to a different step (31, 32, . . . , 39). The targets were numbered serially by step, starting with the first step in the target area. The assignment of targets to states was accomplished on a random basis.

The probabilities of any path intersecting the targets, irrespective of their step location within the target area, are the stationary probabilities of the states. For purposes of evaluating *S*'s performance, the targets were divided into two classes: *high probability targets*, those targets located in the three states having the highest stationary probabilities; *low probability targets*, those targets located in the remaining six states. Given three predictions for each path, *S*'s score for each path was the frequency he predicted High and Low probability targets as the most likely terminal points of the path. Thus, if two of the targets predicted by *S* were in the set defined as High Probability Targets, and one target in the set defined as Low Probability Targets for that path, he was given a score of 2*H*; 1*L*.

Results and Discussion

Path direction was included as an experimental variable to explore the possible influence of this factor on *S*'s performance. The data showed that this variable did not affect the frequency with which High and Low Probability Targets were predicted. Therefore, this variable will not be considered further, and to conserve

TABLE 3

Entropies, $\hat{H} = h_T/N$, of the 10 Sample Paths for the Three Sample Sizes (21, 24, and 27 Steps) and the Entropy of the Information Source

$$(H = -\sum_{i,j} P_i p_{i(j)} \operatorname{Log} p_{i(j)})$$

Path No.	$\hat{H}$ of Paths		
	21 Steps	24 Steps	27 Steps
1	.901	.907	.905
2	.780	.789	.812
3	.859	.865	.874
4	.775	.803	.809
5	.819	.835	.857
6	.895	.952	1.015
7	.841	.843	.862
8	.990	.918	.909
9	.778	.808	.815
10	.891	.885	.901

$$H = \sum_{i,j} P_i p_{i(j)} \log p_{i(j)} = .9628$$

---

[3] The stationary probabilities, P_i, should have been used to select the initial states, since only under these conditions are the stationary probabilities "the probabilities of the system being in any given state at time *t*," i.e., any step *n*; otherwise these probabilities hold only after the system has gone through some large number of steps.

[4] It may be noted that the sample entropies, $\hat{H}$, in Table 3 are not normally distributed about *H*. This result is a consequence of assigning equal probabilities of choice to each of the 21 states when selecting the point of origin for each path, rather than using the stationary probabilities.

space the relevant data showing this result will not be given.

To provide a framework for discussing the variables considered to be significant, the hypotheses that were entertained at the outset of the study will be stated:

Hypothesis 1: Ability of *S*s to predict high probability targets will be greater for low entropy paths than for high entropy paths.

Hypothesis 2: As the length of the path sample size increases, the ability of *S*s to predict high probability targets will be enhanced.

Hypothesis 1 is based on the fact that in paths of a given sample size, low entropy paths, as compared with high entropy paths, provide the *S* with more knowledge about high probability events in the information source. Consequently, the *S*'s ability to select targets located in states with high probabilities of occurrence should be enhanced. Hypothesis 2 is predicted on the proposition that as sample size increases, the *S* receives more knowledge about the probabilities of events emitted by the information source, and therefore, his ability to select targets located in states having a high probability of occurrence will be enhanced.

For purposes of analysis the 10 paths were rank-ordered for each length with respect to their entropy and the 5 paths of lowest entropy were grouped and called "Low Entropy Paths," while the remaining 5 were classed as "High Entropy Paths." To test the hypotheses, chi square tests and differences between percentages were used. Where the chi square test is used, the theoretical expected frequencies of the three high probability targets were defined as 33% of the total number of predictions (three of nine targets), while 67% of the total number of predictions (six of nine targets) was defined as the theoretical expected frequency for the low probability targets. Where differences between percentages are used, the percentages considered were the percentages with which high probability targets were predicted. Table 4 summarizes the data obtained and the results of the chi square tests. These data will be discussed in terms of the hypotheses proposed.

Hypothesis 1.—The column totals in Table 4 contain data relevant to this hypothesis. These data show that for Low Entropy Paths the *S*s predicted high probability targets significantly more frequently than chance ($\chi^2 = 215.6, p < .001$), while for High Entropy Paths the frequency of predicting high probability targets did not depart significantly from chance ($\chi^2 = 2.5$, *ns*). As the data suggest, the percentage with which the *S*s predicted high probability targets on Low Entropy Paths (42.7%) as compared with High Entropy Paths (34.0%) is significantly different ($CR = 9.07$, $p < .001$).

Considering the Entropy columns, a further test of the hypothesis may be made by testing the differences between the percentages with which the *S*s predicted high probability targets on Low and High Entropy Paths for each path length. This analysis produced the following results: at 21 steps—Low = 40.0%, High = 29.4% ($CR = 6.54$, $p < .001$); at 24 steps—Low = 42.5%, High = 34.4% ($CR = 4.87, p < .001$); and at 27 steps—Low = 45.6%, High = 38.2% ($CR = 4.37$, $p < .001$). The results of these analyses are viewed as supporting the hypothesis.

Hypothesis 2.—One set of data relevant to this hypothesis is found in the row totals in Table 4. Here we find that the *S*s predicted high probability targets significantly more frequently than chance at each path length (21 steps—$\chi^2 = 4.5$, $p < .05$; 24 steps—$\chi^2 = 45.9$, $p < .001$; 27 steps—$\chi^2 = 121.2, p < .001$).

The increasing size of the chi square values as the number of steps increases suggests increased ability to make

TABLE 4

FREQUENCY OF PREDICTION OF HIGH AND LOW PROBABILITY TARGETS BY PATH ENTROPY AND PATH LENGTH

Steps	Targets	Low $\hat{H}$ Paths		High $\hat{H}$ Paths		Totals	
		Obs. Freq.	Exp. Freq.	Obs. Freq.	Exp. Freq.	Obs. Freq.	Exp. Freq.
21	High prob.	681	562	495	555	1176	1117
	Low prob.	1023	1142	1188	1128	2211	2270
		$\chi^2 = 37.4$***		$\chi^2 = 9.8$**		$\chi^2 = 4.5$*	
24	High prob.	723	561	594	569	1317	1130
	Low prob.	978	1140	1131	1156	2109	2296
		$\chi^2 = 69.5$***		$\chi^2 = 1.6$		$\chi^2 = 45.9$***	
27	High prob.	770	557	649	560	1419	1117
	Low prob.	919	1132	1049	1138	1968	2270
		$\chi^2 = 121.1$***		$\chi^2 = 20.9$***		$\chi^2 = 121.2$***	
Totals	High prob.	2174	1680	1738	1684		
	Low prob.	2920	3414	3368	3422		
		$\chi^2 = 215.6$***		$\chi^2 = 2.5$			

* $p < .05$.
** $p < .01$.
*** $p < .001$.

accurate judgments. Tests of differences between percentages with which high probability targets are predicted in paths of different lengths produced the following results: 21 steps = 34.7%, 24 steps = 38.4%—$CR = 3.16$, $p < .003$; 21 steps = 34.7%, 27 steps = 41.9%—$CR = 6.10$, $p < .001$; 24 steps = 38.4%, 27 steps = 41.9%—$CR = 2.94$, $p < .01$. These data clearly demonstrate the hypothesized facilitative effect of sample size on *S*s' performances.

Another way of examining the effects of sample size is to consider the data shown in the Steps × Path Entropy cells. Considering first Low Entropy Paths we find that the frequency of predicting high probability targets is significantly greater than chance at each path length (21 steps—$\chi^2 = 37.4$, $p < .001$; 24 steps—$\chi^2 = 69.5$, $p < .001$; 27 steps—$\chi^2 = 121.1$, $p < .001$). Again tests of differences between percentages by path length show that performance improved as sample size increased: 21 steps = 40.0%, 27 steps = 42.5%—$CR = 1.48$, $p < .070$ (one-tailed test); 21 steps = 40.0%, 27 steps = 45.6%—$CR = 3.29$, $p < .001$; 24 steps = 42.5%, 27 steps = 45.6%—$CR = 1.81$, $p < .035$ (one-tailed test).

On the other hand, for High Entropy Paths the frequency of predicting high probability targets was significantly less than chance at 21 steps, $\chi^2 = 9.8$, $p < .01$; at chance at 24 steps, $\chi^2 = 1.6$, *ns*; and significantly greater than chance at 27 steps, $\chi^2 = 20.9$, $p < .001$. These data indicate that sample size has a facilitative effect even in high entropy situations. As might be expected from the preceding data, differences between percentages with which high probability targets are predicted in paths of different lengths show the facilitative effects of sample size (21 steps = 29.4%, 27 steps = 34.4%—$CR = 3.14$, $p < .003$; 21 steps = 29.4%, 27 steps = 38.2%—$CR = 5.43$, $p < .001$; 24 steps = 34.4%, 27 steps = 38.2%—$CR = 2.32$, $p < .02$).

These data clearly indicate that as sample size increases, performance improves. This result is true for both High and Low Entropy Paths. Hypothesis 2 is considered confirmed.

Though the analysis of the data produced statistically significant differences consistent with the hypotheses, psychologically speaking, however, the differences obtained were not of great magnitude. On the other hand, the fact that statistically significant differences in performance were obtained as $\hat{H}$ and sample size, N, were varied does indicate the importance of these variables.

REFERENCES

ATTNEAVE, F. Symmetry, information, and memory for patterns. *Amer. J. Psychol.*, 1955, **68,** 209–222.

ATTNEAVE, F. *Applications of information theory to psychology.* New York: Bolt-Dryden, 1959.

BERLYNE, D. E. Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, **53,** 399–404.

ERIKSEN, C. W., & HAKE, H. W. Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. *USAF WADC tech. Rep.*, 1954, No. 54–162.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., & LEONARD, A. J. Stimulus correlates of visual pattern recognition. *J. exp. Psychol.*, 1956, **51,** 1–11.

HAKE, H. W., & GARNER, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. exp. Psychol.*, 1951, **42,** 358–366.

KEMENY, J. G., SNELL, J. L., & THOMPSON, G. L. *Introduction to finite mathematics.* Englewood Cliffs, N. J.: Prentice-Hall, 1957.

KLEMMER, E. T., & FRICK, F. C. Assimilation of information from lot and matrix patterns. *J. exp. Psychol.*, 1953, **45,** 15–19.

MILLER, G. A., & MADOW, W. G. On the maximum likelihood estimate of the Shannon-Weaver measure of information. *USAF Cambridge Res. Cent. tech. Rep.*, 1954, No. 54–73.

MOOD, A. F. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

SHANNON, C. E., & WEAVER, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

(Received June 1, 1962)

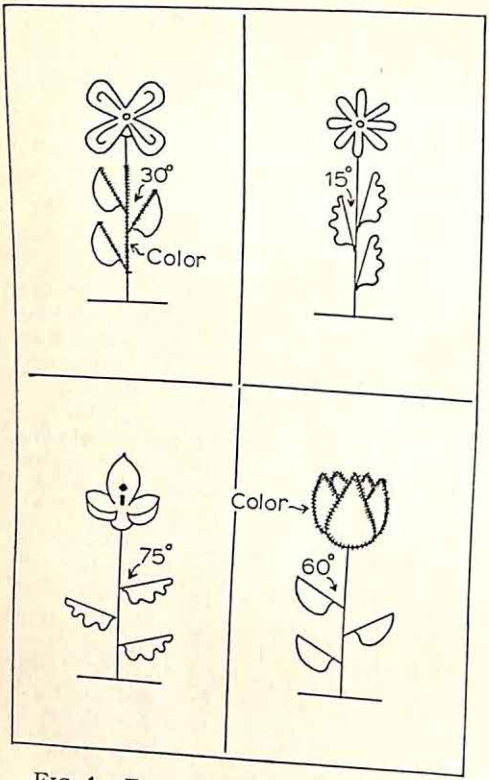


FIG. 1. Examples of flower patterns used in experimental problems. (The degrees indicate the size of the angle of the leaves to the stem. The cross-hatched portions represent areas which were colored to obtain emphasis or counteremphasis.)

The values of the stimulus dimensions listed in Table 1 were as follows: Angle of the leaves to the stem (30° or 60° when relevant; 15° or 75° for Group T2; 45° when constant); Flower (tulip, daisy, pansy, or fleur-de-lis); Leaf Position (2 left-1 right or 2 right-1 left); Leaf Shape (smooth or serrated); Color (red or green).

Group C, the principal control group, practiced only on the basic "hard" problem. Nine other groups learned Problems T1–T9 and then transferred to the problem of Group C. Both their training and transfer performances were then compared to the performance of Group C.

In the problems presented to Groups T1–T4, the relevant angle dimension was emphasized in one of three ways: (*a*) For Group T1, the irrelevant dimensions were fixed so that only the relevant angle varied during training; (*b*) For Group T2, the angle difference was doubled; (*c*) For Groups T3 and T4, color was constant and spatially contiguous to the angle dimension. Groups T5–T7 had problems where the emphasizer role of color was varied. For Group T5, color was an emphasizer but irrelevant. For Group T6, color was an emphasizer but relevant. For Group T7, color was a counteremphasizer, located on the flower, and was relevant. Groups T8 and T9 were control groups for color. For T8, color was relevant and located on the fixed angle. For T9, color was relevant and located on the flower; the angle was fixed.

Subjects.—The *S*s were 215 elementary psychology students at Michigan State University who were randomly assigned to 10 groups, each composed of 20 *S*s. Fifteen *S*s failed to solve their training problems and were replaced by *S*s who did.

RESULTS AND DISCUSSION

The groups did not differ on the practice problem, F (9, 205) = 1.29, $p > .05$, and there were no significant correlations between performance on the practice and training problems.

In Table 2, the mean performance of each training and transfer condition is compared with that of Group C. For training comparisons, the data are total errors made by the first 20 *S*s (solvers and nonsolvers) in each group. For transfer, the data are total errors made by the 20 *S*s who solved and were transferred in each group.

Training.—The three basic emphasizer groups, T1–T3, learned significantly faster than Group C. Removing the irrelevant dimensions for Group T1 led to the fastest acquisition; doubling the angle difference for Group T2 and coloring the angle red for Group T3 reduced errors by one-half and one-third, respectively. Green had no emphasizer effect for Group T4, possibly because green is strongly associated with flower stems whereas red is not.

Group T5 learned slightly but not significantly faster than Group C. Generally, learning is retarded by

TABLE 1

Experimental Problems

Group	Stimulus Description			Nature of Emphasis
	Angle	Color	Flower, Leaf Position, and Leaf Shape	
C	Relevant	Absent	Irrelevant	None
T1	Relevant	Absent	Constant	Irrelevant cues fixed
T2	Relevant	Absent	Irrelevant	Angle difference doubled
T3	Relevant	Constant	Irrelevant	Angle colored red
T4	Relevant	Constant	Irrelevant	Angle colored green
T5	Relevant	Irrelevant	Irrelevant	Angle colored red or green but color uncorrelated with angle
T6	Relevant	Relevant	Irrelevant	Angle colored red or green but color correlated with angle
T7	Relevant	Relevant	Irrelevant	Flower colored red or green but color correlated with angle
T8	Constant	Relevant	Irrelevant	Angle fixed and colored red or green
T9	Constant	Relevant	Irrelevant	Angle fixed; flower colored red or green

added irrelevant cues so that the observed facilitation indicates that the emphasizer role of color predominated. Thus, it seems possible to add an irrelevant dimension which serves as an emphasizer and hence facilitate learning.

Color was more salient as a cue than angle. Both Groups T8 and T9 made significantly fewer errors than Group C. Angle and color cues seem additive since Groups T6 and T7 made numerically fewer errors than their controls, Groups T8 and T9, which had only color relevant or Group C which had only angle relevant.

Transfer.—Groups T1–T5 showed nearly perfect transfer to the problem of Group C. Thus, an emphasizer which could not serve as the basis for a correct solution had no disruptive effect when removed on transfer.

Training *S*s on an easy problem with redundant and relevant cues was not efficient. Apparently, most

TABLE 2

Means and *SD*s of Errors in Training and Transfer

Group	Training			Transfer		
	Mean	*SD*	*t*[a]	Mean	*SD*	*t*[a]
C	19.50	10.51	—	—	—	—
T1	2.60	2.06	7.07	0.40	1.35	—[b]
T2	8.35	9.07	3.60	0.20	0.62	—[b]
T3	12.45	11.79	2.00	0.00	0.00	—[b]
T4	19.40	13.01	0.03	0.05	0.21	—[b]
T5	14.65	12.99	1.31	0.00	0.00	—[b]
T6	3.45	5.31	6.10	16.45	14.45	0.76
T7	2.40	3.59	6.90	23.45	12.14	1.13
T8	4.05	8.49	5.13	20.10	13.86	0.15
T9	3.40	5.31	6.12	19.75	14.01	0.06

[a] As compared with Group C, $t_{.05} = 2.04$, $df = 30$.

[b] Statistical test unnecessary.

*S*s in T6 and T7 solved their training problem on the basis of color since these groups showed no significant transfer to the Group C problem and the control groups, T8 and T9, did not differ on transfer from Group C. For Group T6, 5 *S*s showed perfect transfer; only 1 *S* did so in Group T7.

To summarize, these results extend Lawrence's (1952) finding to two-choice concept identification. Training on an easy problem and transfer to a hard one with the same relevant cue is more efficient than training on the hard one alone, provided that alternative solutions are not present during training. The data also suggest that the degree of efficiency depends upon the kind of stimulus emphasis employed.

A logical distinction between a stimulus' saliency as a cue and its saliency as an emphasizer should be made. To be a cue, a stimulus has to have at least two values, whereas an emphasizer may have only one. The weight of a cue can be reduced or enlarged by changing the discriminable difference between values but its saliency as an emphasizer may be unchanged. Group T3 is a case where color was constant, i.e., had no cue value, but an emphasizer effect was observed.

An emphasizer would seem to have maximal effect when the relevant cue is embedded and dissimilar to other cues. A reconsideration of the present problem suggests that the flower patterns, with distinct but similar angles, were not optimal configurations for finding a strong emphasizer effect with color. The Group C problem was difficult both because the angles were similar and irrelevant stimuli were present. A larger emphasizer effect was obtained by either removing irrelevant cues or doubling the angle difference than by color emphasis. Hull's (1920) procedure of coloring the embedded radicals in the Chinese characters apparently provided more opportunity for an emphasis effect by color contrast. Apparently, an emphasizer has its effect on the probability of *S* attending to a dimension and not on discriminating one value from another.

Presolution analysis.—The presolution phase is the sequence of errors and successes made by *S* prior to his last error. To compare incremental with all-or-none theory, Monte Carlo computations were made on theoretical learning curves exactly fitting the Bourne-Restle (1959) model for values of $\Theta \geq .01$, estimated individually for 222 "Stat-*S*s." The Stat-*S*s correspond directly with *S*s in the experiment and were "run" to the same criterion of learning.

The hypothesis that the probability of an error was stationary was tested by comparing error frequencies between the first and second halves of the presolution phase for each *S*. As expected from all-or-none theory, for the 176 *S*s who solved and made 2 or more errors, the averages per half were nearly equal: 4.90 and 5.12 mean errors, respectively, t (175) $= 1.32$, $p > .18$. First half errors were more frequent for the Stat-*S*s, averaging 7.26 and 5.16 errors, t (147) $= 9.22$, $p < .01$.

Stationarity was further tested in two ways. First, the probability of a success, conditional on *S* being in the presolution phase, is plotted for solvers against trials (for five trial blocks) in Fig. 2a.

The data in Fig. 2a are stationary over some 60 trials. Using a chi square test (Suppes & Ginsberg, 1961), the hypothesis of stationarity was not rejected, χ^2 (11) $= 10.71$, $p > .30$; but was rejected for the Stat-*S*s, χ^2 (11) $= 38.88$, $p < .01$. To test for trend, Kendall tau coefficients (Siegel, 1956) were computed between the p values and trials. No significant trend was noted in the data ($\tau = -.17$, $z = 0.77$, $p > .44$) while the Stat-*S*s conditional probabilities tended to increase ($\tau = +.33$, $z = 1.50$, $p > .13$).

Second, backwards learning curves

TABLE 3

TRANSITION FREQUENCIES FOR INDEPENDENCE TEST: FREQUENCY OF SUCCESS OR FAILURE ON TRIAL $n + 1$ GIVEN A SUCCESS OR FAILURE ON TRIAL n

	Trial $n+1$		Conditional Probability of Success on Trial $n+1$
	Success	Fail	
Trial n			
Success	1575	1594	.497
Fail	1614	1697	.487

(Hayes, 1953) obtained for 220 solvers and the Stat-*S*s are shown in Fig. 2b. The hypothesis of stationarity was not rejected on the data, χ^2 (23) = 21.66, $p > .50$; but was rejected on the Stat-*S*s, χ^2 (23) = 133.55, $p < .01$. No significant trend was found in the data ($\tau = .17$, $z = 1.19$, $p > .23$) but it was significantly decreasing for the Stat-*S*s ($\tau = .74$, $z = 5.09$, $p < .01$). These results clearly favor the all-or-none interpretation.

Independence of trials was tested on the data by comparing frequencies of a success or an error on Trial $n+1$ following a success or an error on Trial n during presolution for all *S*s. Chi square tests (Suppes & Ginsberg, 1961) were not significant for 10 training and four transfer groups. The combined data from all groups are given in Table 3.

The transition frequencies in Table 3 were independent, χ^2 (1) = 0.59, $p > .30$.

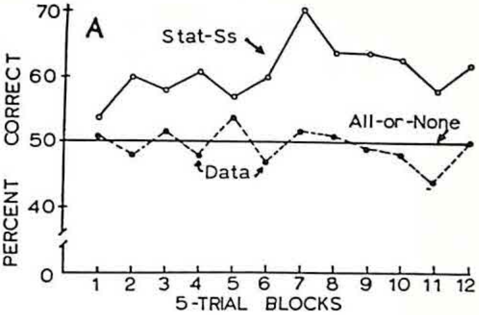
Error distributions.—The theoretical distribution of total errors to solution is geometric (Restle, 1961, 1962). Nonsolvers, in the present study, should produce a binomial distribution with probability of an error near one-half for 64 trials. The complete distribution is thus a combination of geometric distribution (for solvers) and a binomial distribution (for nonsolvers). Estimates of c, the learning parameter, must take into account both types of *S*s. If N_s *S*s solve and there are T total errors made by a group, then the maximum likelihood estimate of c and its variance can be shown to be (Trabasso, 1961)

$$\hat{c} = N_s/N_s + T \qquad [1]$$

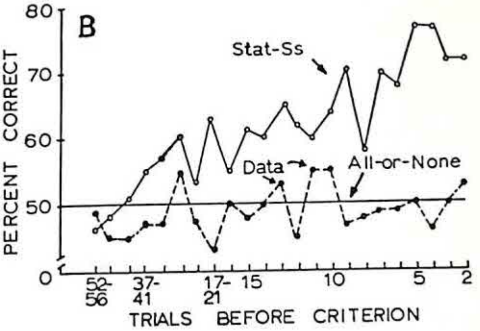
and

$$\text{var}\ (\hat{c}) = \frac{\hat{c}^2(1 - \hat{c}).}{N_s}$$

Estimates of c were obtained for 10 training and 2 transfer groups, T8 and T9. Four groups, the training of C and T4 and the transfer of T8 and T9, totaling 86 *S*s, gave identical estimates of $c = .035$. For comparison, the learning rate (θ) of these 4 groups was estimated from the Bourne-Restle model and found to be .067. Monte Carlo computations were made on a theoretical



2a. Percentage of correct responses conditional on *S* being in the presolution phase.



2b. Backwards learning curve showing percentage of correct responses on trials prior to criterion.

FIG. 2. Stationarity curves. (The data points are from *S*s who solved in either training and/or transfer.)

$(A)/\text{rm}(A)+\text{m}(C_a) = .201$. The remaining 79.9% of the *S*s should learn with $c = .035$. Thus $20 \times .237 = 4.72$ *S*s should make no errors; five were observed. Expected errors (using Group T8's control data) were $20.10 \times .799 = 16.07$, not significantly different from the 16.45 observed ($t = 0.12$).

Similar calculations hold for Group T7. On transfer, 1.66 *S*s were expected to show perfect transfer; 1 did. Predicted mean errors (using Group T9 as a control) were 18.11, significantly different from the 23.60 observed ($t = 2.09$). Thus, whereas slight positive transfer was expected, mild negative transfer was observed. The result may be due to sampling variability (note large σ's in Table 2) or to *S*s in T7 continuing to look at the location (flowers) which previously bore the relevant colors (now absent). The second possibility seems ruled out by the transfer result of Group T9.

Since one transfer prediction is in error, these results do not conclusively favor Restle's theory. They do, however, indicate the usefulness of the theory as a baseline for evaluation of transfer.

The data from presolution and error distribution analyses clearly indicate an all-or-none or one-step process in single classification concept learning. This is not to say that all concept learning can be characterized by a one-step process. It is possible that an increase in problem complexity, such as learning conjunctive or relational concepts, would lead to learning which more resembles an incremental process. However, at present, it is not clear whether learning in such cases would be in fact incremental or a compounding of possibly two or more all-or-none processes (cf. Theois, 1963). The issue is still to be decided.

References

Baker, R. A., & Osgood, S. W. Discrimination transfer along a pitch continuum. *J. exp. Psychol.*, 1954, **48**, 241–246.

Bourne, L. E., Jr., & Restle, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

Bower, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255–280.

Estes, W. K. Learning theory and the new mental chemistry. *Psychol. Rev.*, 1960, **67**, 207–223.

Hayes, K. J. The backward curve: A method for the study of learning. *Psychol. Rev.*, 1953, **60**, 269–275.

House, B. J., & Zeaman, D. Transfer of a discrimination from objects to patterns. *J. exp. Psychol.*, 1960, **59**, 298–302.

Hovland, C. I. A set of flower designs for experiments in concept formation. *Amer. J. Psychol.*, 1953, **66**, 140–142.

Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**(1, Whole No. 123).

Lawrence, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

North, A. J. Acquired distinctiveness of form stimuli. *J. comp. physiol. Psychol.*, 1959, **52**, 339–341.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Restle, F. Additivity of cues and transfer in discrimination of consonant clusters. *J. exp. Psychol.*, 1959, **57**, 9–14.

Restle, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, **26**, 291–306.

Restle, F. The selection of strategies in cue learning. *Psychol. Rev.*, 1962, **69**, 329–343.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

Theios, J. Simple conditioning as two-stage all-or-none learning. *Psychol. Rev.*, 1963, **70**, in press.

Trabasso, T. R. Additivity of cues in discrimination learning of letter patterns. *J. exp. Psychol.*, 1960, **60**, 83–88.

Trabasso, T. R. The effect of stimulus emphasis on strategy selection in the acquisition and transfer of concepts. Unpublished doctoral dissertation. Michigan State University, 1961.

Suppes, P., & Ginsberg, R. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. Psychological Series Technical Report No. 39, 1961, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Warren, J. M. Additivity of cues in visual pattern discriminations by monkeys. *J. comp. physiol. Psychol.*, 1954, **47**, 290–292.

(Received June 4, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 407–413

PERCEPTION AND SHORT-TERM MEMORY IN THE IDENTIFICATION OF MULTISYMBOL DISPLAYS[1]

WARREN H. TEICHNER

University of Massachusetts

It has been suggested that a direct estimate of the amount of immediate forgetting in responding to multisymbol displays may be obtained as the difference between the spans of perception (a count of the symbols) and of short-term memory (number of named symbols). An experiment was designed to determine whether the procedure used to obtain this difference produced an artifact due to differences in sets rather than a memory measure. The results failed to support this possibility. They did suggest that the accuracy of naming is increased when S both counts and names. In addition, the results, extending those of previous studies, suggested that both spans may become asymptotic with slightly more than 9 displayed symbols.

Miller (1956) has reviewed evidence supporting the notion that the basic low capacity of man to process information is the result of limitations in perceptual spans and in the span of short-term memory. This two-process distinction has had long acceptance (Woodworth & Schlosberg, 1954), but until recently (Sperling, 1960; Teichner, Reilly, & Sadler, 1961) no effort seems to have been made to establish the distinction in terms of a set of experimental operations.

The procedure suggested by Teichner, Reilly, and Sadler (1961) and Teichner and Sadler (1962) for distinguishing between perception and short-term memory makes use of two independent *S* groups, each with different instructions. One group is required to *name* the different letters (categories) on briefly-exposed multiletter displays without regard to the density of repetition of the letters (naming group); the other is required to report the *number* of different categories, also without regard to their density (counting group). The perceptual process is assumed to be identical for the two groups. That is, since both groups must report letters without regard to density, both groups must discriminate each displayed letter in order to reject letter repetitions. Thus, the perceptual task is assumed to be the same whether the letters are to be counted or named. Similarly, the requirement for recognition of what has been seen is assumed to be the same during the exposure period. The two groups are assumed to differ primarily in regard to their reproductive memory load, i.e., after the display has been removed one must recall only a digit whereas the other must recall the names of the letters.

Rather than represent a perception-memory distinction, differences in the number of categories reported by the two groups may be due simply to differences between a counting set as compared to a naming set. If so, it is not unreasonable that more letters might be reported with a counting set than with a naming set.[2] Other set variables may also be involved. For example, in the previously cited experiments *S*s were given no instructions one way or the other about guessing. If they had been encouraged to guess or provided with a guessing penalty, they might have reported different numbers of categories. It appears important, therefore, that, before continuing use is made of the perception-memory distinction, these set variables be investigated. This was one purpose of this experiment. To achieve it, comparisons were made between number of named and counted letters when *S*s were required both to name and count and when they did only one of these. The guessing set variable was also studied. In addition, this experiment was used to extend study of the effects of number of categories and of density by extending the range of the former over that used in previous experiments.

---

[1] This research was carried out under Contract N61339-588, with the United States Naval Training Device Center, Port Washington, N. Y.

Benjamin Naylor assisted in this experiment.

METHOD

Apparatus.—The apparatus was essentially that used by Teichner, Reilly, and Sadler (1961). One hundred and fifty slides were used. These varied in number of categories from 5 to 9. Density, defined as the frequency of occurrence of each category, was 1, 3, or 5 and was constant for all categories on a given slide. Thus, there were 15 density-category combinations. Each of these combinations was replicated 10 times with a unique random sample and random letter locations to make up the 150 slides. Exposure time was 1 sec.; interslide interval was 10 sec.

Procedure.—Forty-eight *S*s were assigned randomly to eight groups of 6 *S*s each. Each group was run as a whole through the entire 150 slides in a single experimental session. The slides were presented in a random order which was the same for each group. Prior to the experimental slides, 2 sample slides were presented for practice. Using a prepared data sheet, *S* recorded his own response during the interslide interval.

Prepared instructions were read to each group. Four groups were instructed to be as accurate as possible in their reports. These groups were also informed that wrong responses would not be counted against them and they were urged to guess if they felt any uncertainty. Four other groups were also instructed to be as accurate as possible, but were told that wrong responses would be counted heavily against them and that they should not guess at all.

The guess and no-guess groups received the same instructions for counting and naming. Within each of these two major groups there was a group which reported only the names of the letters. Similarly, there were two groups which reported only the number of letter categories. In addition, there were two other subgroups within each of the guess and no-guess main groups: (*a*) a group which recorded the names of the letters and then immediately reported the number of letters and (*b*) a group which first reported the number of letters and then their names. The *S*s were never informed about the range of categories and densities nor were they given knowledge of their performance.

TABLE 1

ANALYSIS OF VARIANCE OF NUMBER OF LETTERS REPORTED BY THE COUNTING GROUPS IN EXP. I

Source	*df*	*MS*	*F*
Between *S*s total (*S*)	35		
Guess vs. No-Guess (G)	1	4.440	0.294
Order (N)	2	26.410	1.750
G × N	2	25.740	1.705
S/NG	30	15.089	
Within-*S*s total	504		
Density (D)	2	1.925	1.824
Category (C)	4	6.747	36.470***
D × C	8	0.326	4.233***
G × D	2	1.975	1.872
G × C	4	0.065	0.351
D × N	4	1.682	1.594
C × N	8	0.245	1.324
G × D × C	8	0.391	5.077***
C × N × G	18	0.195	1.054
D × C × N	16	0.131	1.701
D × N × G	4	2.370	2.246
D × C × N × G	16	0.292	3.792***
S × D/NG	60	1.055	
S × C/NG	120	0.185	
S × D × C/NG	240	0.077	

*** $p < .001$.

---

[2] First suggested to the writer by Douglas Ellson (personal communication).

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF LETTERS REPORTED BY THE NAMING GROUPS IN EXP. I

Source	*df*	*MS*	*F*
Between-*S*s total (*S*)	35		
Guess vs. No-Guess (G)	1	.0327	1.466
Order (N)	2	.0740	3.318*
G × N	2	.0204	0.915
S × G × N	30	.0223	
Within-*S*s total	504		
Category (C)	4	.0432	54.000***
Density (D)	2	.0050	5.000**
C × D	8	.0022	3.667**
G × C	4	.0000	0.000
G × D	2	.0008	0.800
N × C	8	.0008	1.000
N × D	4	.0024	2.400
G × C × D	8	.0011	1.833
N × C × D	16	.0004	0.667
G × N × C	8	.0015	1.875
G × N × D	4	.0017	1.700
G × N × C × D	16	.0009	1.500
S × C/GN	120	.0008	
S × D/GN	60	.0010	
S × C × D/GN	240	.0006	

* $p = .05$.
** $p < .01$.
*** $p < .001$.

Subjects.—The *S*s were undergraduate students from an introductory psychology class. No *S* had participated in any previous experiments. Each group of 6 *S*s competed among its own members for a cash prize of $3.00.

RESULTS

Because of the complex nature of between- and within-group comparisons, an analysis of variance, summarized in Table 1, was performed on the number of letters reported by the counting groups and another analysis of variance, summarized in Table 2, was performed on the number of letters reported by the naming groups. In neither case was there a significant difference between guessing and not guessing, even though both the number counted and the number named were consistently, though only slightly greater, with guessing as compared to not guessing.

The order effect (i.e., counting first or naming first) was not significant for the counting group and was just significant at the .05 level for the naming group. For the naming group both categories and density were significant at least at the .01 level as was the interaction of these two variables. For the counting group, categories and the interaction of categories with density was significant at or beyond the .01 level, but the main effect of density was not significant. Thus, these analyses suggest that the set variables had no effect at all on the performance of the groups, but that the order of counting and naming may have had an effect on the number of letters named.

Figure 1 was prepared to examine the order effect more closely and to compare the general differences between naming and counting with regard to the number of letters reported. The data presented are averaged over guessing conditions and densities. The straight line represents perfect performance; i.e., this line represents the result that would have been obtained had the *S*s reported exactly the number of letter categories presented. The results are clear in showing that both naming and counting produced an increasing underestimate and that in both cases the trend of number of letters reported was increasing but at a very slow rate. With regard to the different orders of naming, it can be seen that accuracy was greatest in order of naming first, last, and only. These differences which are just significant at the .05 level are very small and all of less accuracy than the number of letters counted, which is presented as the mean value across all three counting orders. To test the hypothesis that the naming data were less accurate than the counting data, as shown in Fig. 1, *t* tests of the overall differences were performed for each order condition. That is, the differ-

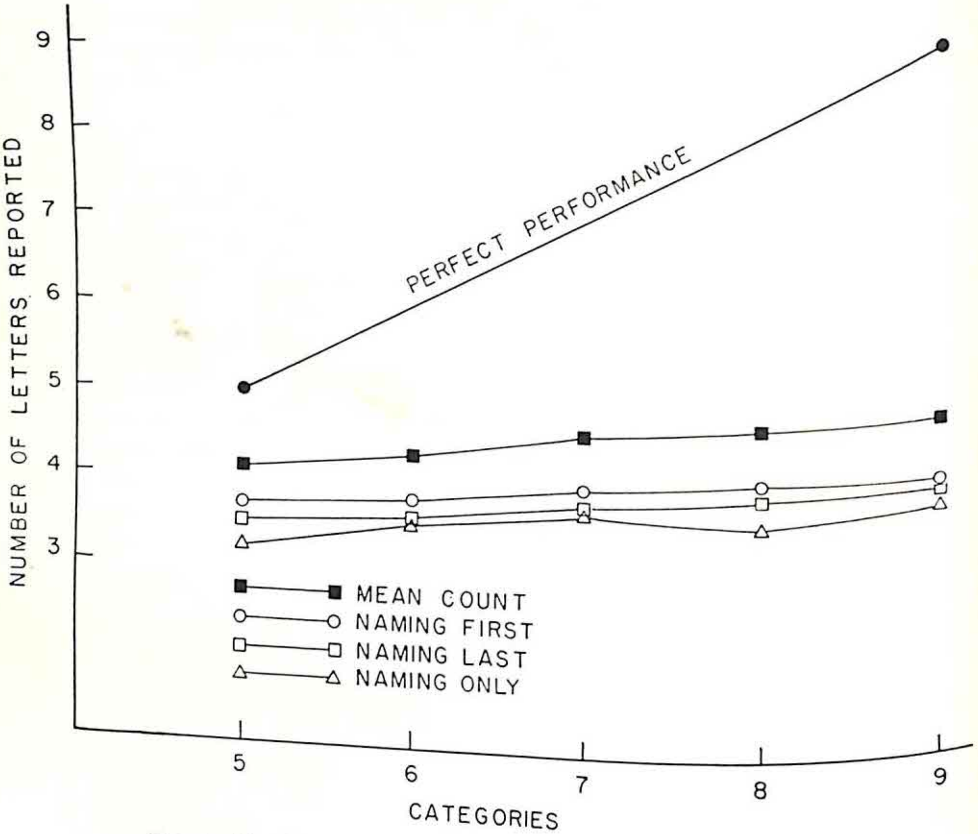


FIG. 1. Number of letters named with three naming conditions compared to the number of letters counted.

ence between naming only and counting only was evaluated as were the differences between naming first and counting first and naming second and counting second. In the first case, p was less than .005, in the second case less than .05, and in the third case less than .0001. Thus, it seems clear that the differences between counting and naming represented by the differences between the mean counting line of Fig. 1 and each naming line were not random effects.

Figure 2 shows the average effects of letter categories and densities on the number of letters named (Fig. 2a) and on the percentage of correctly named letters (Fig. 2b). It will be recalled that for the number of letters counted density was not a significant effect. Figure 2 suggests that the barely significant density effect obtained for naming should not be given much weight. For both of the measures shown the differences among densities were small and not systematic. On the other hand, the number of letters named tended to increase slightly with an increase in the number of categories; whereas the percentage of letters correctly named decreased in a fairly steep, linear fashion.

Figure 3 shows the effects of order on accuracy. As with Fig. 2b, this figure suggests an essentially linear decline of accuracy with increasing numbers of categories. This figure also shows that the requirement to count as well as to name aided accuracy of naming since the poorest performance was obtained with the condition in which *S*s only named the letters.

Discussion

The results suggest that of the set factors studied only the effects of order of naming or counting on the number of letters named might have an effect on the perception vs. short-term memory distinction. This effect was barely significant at the .05 level; it was not large nor did it have a substantial effect in reducing the differences between counting and naming. Interestingly, this result suggests that when *S* is required to count as well as to name, his naming performance is aided and this is true both in terms of the number of categories reported and their accuracy. The reverse effect was not true, i.e., being required to name the letters had no effect

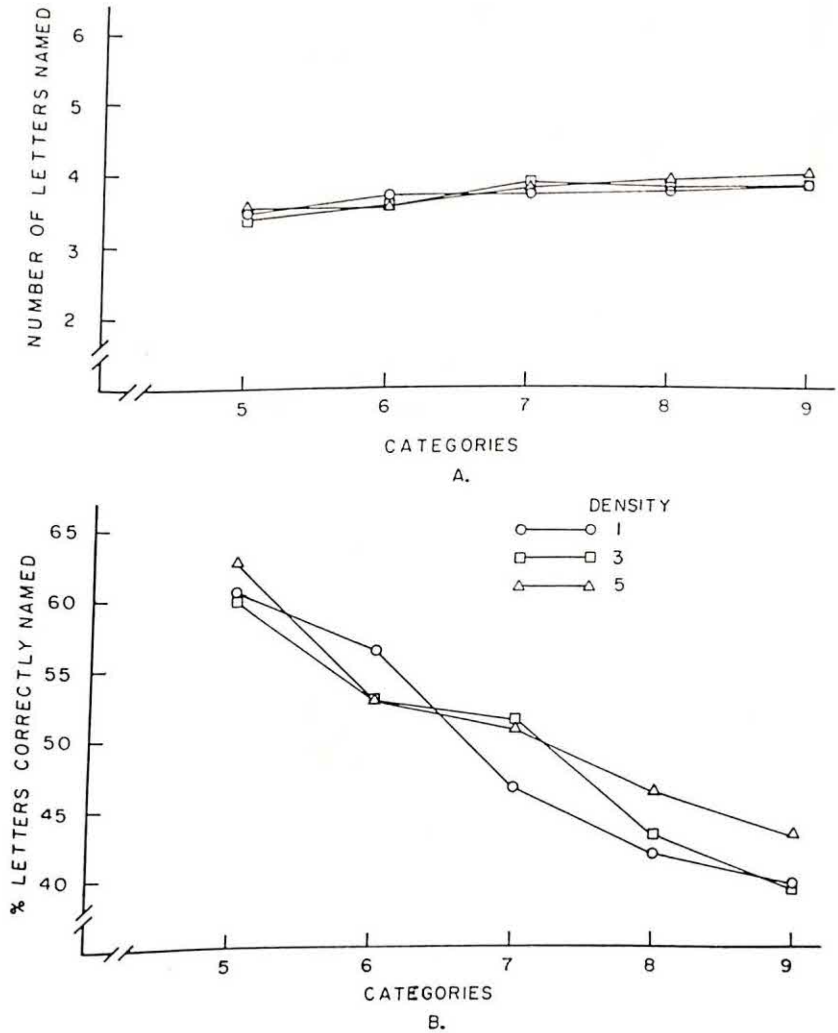


FIG. 2. Effects of density and categories on the number of letters named and the accuracy of naming.

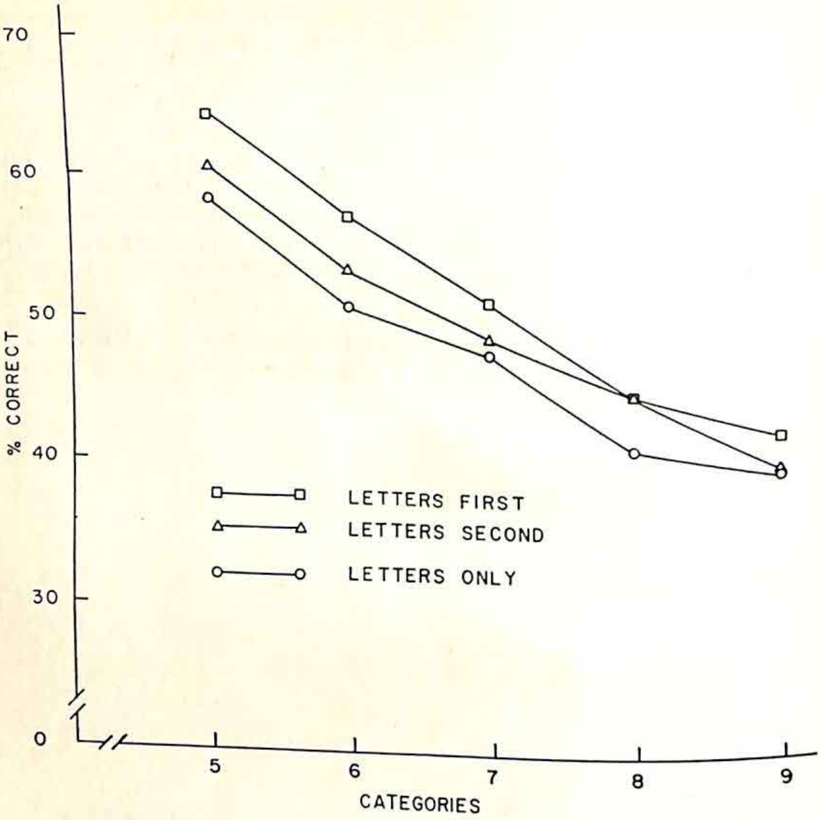


FIG. 3. Accuracy of naming with three naming conditions as a function of categories.

on counting. As a hypothesis, it is possible that the slight additional memory load represented by the count of the letters was more than offset by the count which may have provided *S* with a standard for evaluating his own recall and, thereby, modifying it. In any case, the effect is small, and particularly in view of the finding that it *aids* rather than hinders naming performance, it does not seem to provide a basis for rejecting the procedure of taking the difference between naming and counting as short-term forgetting. It does suggest that the absolute size of the difference will be affected slightly if *S* must do both.

Previous studies (Teichner, et al., 1961) have shown that for a 1-sec. exposure time, as used here, density has no effect on performance. The present results tend to bear this out again, although the density effect for naming was significant at the .05 level. In view of the smallness of this effect, its nonsystematic nature and the results of previous studies, the present significance of density would not appear to warrant serious consideration.

This study differed from previous ones in that in this the letter categories ranged from 5 to 9, rather than from 1 to 6. In the previous studies it was found that the accuracy of naming performance was inversely proportional to the number of categories, while the number of letters counted and named first overestimated and then underestimated perfect performance or perfect information transmission. The present results are consistent in showing that as the number of categories is increased up to nine, the size of the underestimate increases. Further, inspection of the data suggests that the number of letters reported appears to be approaching an asymptote somewhere shortly beyond nine letters. The increasing underestimate, then, appears to be the result of a failure of *S* to keep increasing his estimate as the amount of information presented increases rather than simply a failure to increase it enough.

REFERENCES

MILLER, G. A. The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

SPERLING, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74** (11, Whole No. 498).

TEICHNER, W. H., REILLY, R., & SADLER, E. Effects of density on identification and discrimination in visual symbol perception. *J. exp. Psychol.*, 1961, **61**, 494–500.

TEICHNER, W. H., & SADLER, E. Effects of exposure time and density on visual symbol identification. *J. exp. Psychol.*, 1962, **63**, 376–380.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received June 11, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 414–419

A COMPARISON BETWEEN CORRECTION AND NONCORRECTION METHODS IN DRIVE DISCRIMINATION[1]

NORMA F. BESCH, HERMAN MORRIS, AND SEYMOUR LEVINE[2]

Columbus Psychiatric Institute and Hospital, Ohio State University

The hypothesis that the use of the correction method in drive discrimination experiments contributes to the slow learning found in these situations was supported. A T maze was used for the training of 30 rats in a hunger-thirst discrimination for a social reward. No learning was found for the Ss of the Correction group, which were allowed to retrace to the reward side after an incorrect choice. Only the Ss of the Noncorrection Alley group (NC-A), which were confined in the incorrect arm of the maze after an incorrect choice, performed significantly better than chance in the last half of the 60-day learning period. Both the Ss of the Noncorrection Mesh group, which were detained in a mesh enclosure attached to the incorrect arm, and the Ss of the NC-A group performed significantly better than chance in the last 4 wk. of the learning period.

Since the cue properties of drive are extensively used theoretically, it is of importance that drive discriminations are not easily established. Therefore the possibility that the procedures used in drive discrimination experiments are nonoptimum should be investigated. The present experiment was designed to determine if the correction method, a standard procedure in drive discrimination experiments, contributes to the difficulty in establishing learning in these situations. In a review of the drive discrimination literature, Webb (1955) suggested that "the extent to which drives serve as cues is a decreasing function of the availability of systematic externally generated cues." Kalish (1946) demonstrated the superiority of the noncorrection over the correction technique in a spatial learning situation in which differential external cues were absent. In drive discrimination learning situations where differential external cues are held at a minimum, it seems likely that the correction method may be inferior to the noncorrection method. The only evidence consistent with this hypothesis is not conclusive. Using a noncorrection method, Leeper (1935) found considerably faster learning than Hull (1933), who had used the correction method in a similar experimental arrangement.

METHOD

Subjects.—Thirty male rats of the Sprague-Dawley strain from 90–110 days old were used as *Ss*. They were assigned 10 each to one of three groups designated as Correction (C), Noncorrection Mesh (NC-M), and Noncorrection Alley (NC-A).

Apparatus.—The floor plan of the apparatus which was 7 in. in depth is shown in Fig. 1. The starting section, stem, and arms of the pine T maze were painted flat black. The floors of these sections were covered with rubber matting; the top was covered with hinged pieces of clear Plexiglas. The goal boxes were constructed of black Plexiglas; the tops were hinged for easy removal of *Ss*. The mesh sections were made of steel; hinged doors on the tops of these sections were used for removal of *Ss*. The partitions in the goal boxes were of clear Plexiglas and could be removed or locked in place from outside the

---

[1] This research was supported in part by Grant 04456 from the National Science Foundation.

[2] Now at Department of Psychiatry, Stanford University.

goal boxes. The partitions were moved through narrow cutouts in the walls of the goal box. One-way doors, 4 in. high and of black opaque Plexiglas, served as the entrances to the goal boxes; these doors could be locked from outside the goal box by the *E*. All other doors, which were guillotine doors of black Plexiglas, were operated from behind the start box, by means of a pulley system made possible by steel crosspieces above the maze.

Illumination was provided by a partially blackened 15-w. bulb placed behind the choice point and approximately 6 in. above the top surface of the maze. The maze was rotated approximately 45° in a clockwise or a counterclockwise direction randomly over days.

Model 330S of the Photo Contact Relay, Hunter Manufacturing Company, Incorporated, was used in the timing circuit. The light was provided by three well-shielded 3-w. lamps mounted on the outside of the apparatus directly opposite the units. The light passed through round holes, .5 in. in diameter, which were covered by Wratten filters, No. 88A. When an *S* interrupted the infrared beam activating a photocell, 3 in. from the start-box door, a Standard Electric clock started. The clock was stopped by the interruption of one of the beams activating one of two photocells, each located between the choice point and a guillotine door, 1 in. from the guillotine door.

Procedure.—The *S*s were given 10 days habituation to the hunger and thirst schedules which were used during the experimental phase. Hunger and thirst days were assigned randomly for the 10-day Habituation Period and the 69-day Experimental Period, with the restriction that no more than 3 hunger or thirst days would occur in succession. The daily Habituation and Experimental Periods consisted of three successive 2-hr. periods. The second 2-hr. period was reserved for the experimental treatment; no food or water was present during this period. On hunger days, during the first 2-hr. period, *S*s were given free access to water in individual feeding and drinking compartments. During the third 2-hr. period, they were fed their daily ration of 11 gm. of Purina kibbled meal; free access to water was given in the last hour. On thirst days *S*s were given 5.5 gm. of meal at the beginning of each hour of the first period; free access to water was given during the third period. Thus the hours of deprivation at the time of testing were 20 hr. for water and 20 or 24 hr. for food, on thirst and hunger days, respectively. Every seventh day of the Experimental Period during the last hour of the third 2-hr. period, *S*s were given access to an unlimited amount of food with water present. No experimental treatment was given on the following day; therefore, in the 69-day Experimental Period there were only 60 experimental treatment days.

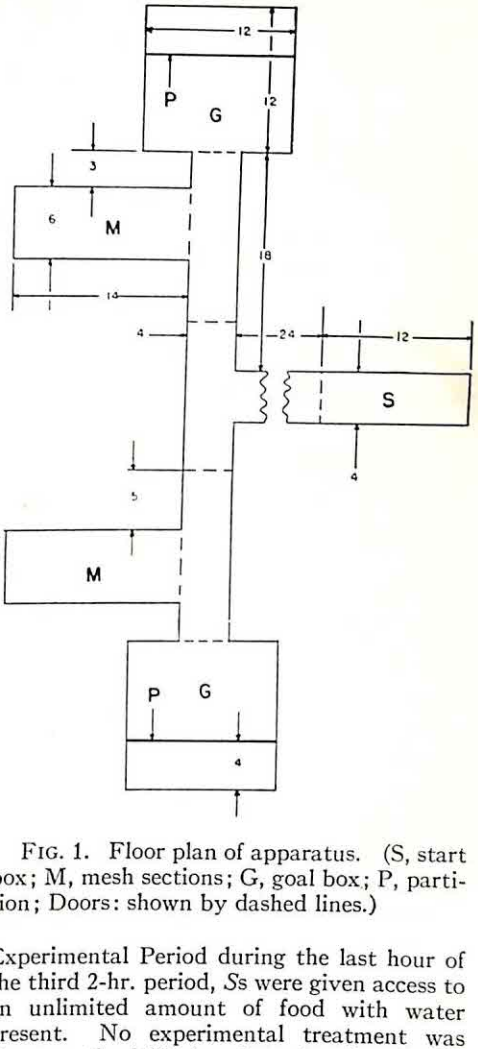


FIG. 1. Floor plan of apparatus. (S, start box; M, mesh sections; G, goal box; P, partition; Doors: shown by dashed lines.)

In addition to the ad lib. feeding every seventh day, *S*s were fed fruit three times a week, immediately after being returned to their home cages. On the regimen used, *S*s were at 95 and 100% original body weight, respectively, on the last day of the Habituation Period and the last day of the Experimental Period.

On the final day of the Habituation Period, *S*s were placed in the apparatus and allowed to explore freely for 5 min. The initial choice of a right or left turn was recorded. Within the categories of right and left turners *S*s were assigned randomly to experimental groups. Three trials were given on Day 1 of the Experimental Period; six trials a day were given thereafter. The running order for the *S*s was assigned randomly at the beginning of the experiment and remained constant thereafter. The intertrial interval was approximately 10 min. The *S*s were housed in group cages until Day 52 of the Experimental Period or Treatment Day 46, after which they were housed in individual cages.

The discrimination problem consisted of turning right when hungry and left when thirsty for a social reinforcement. In view of Webb's (1955) suggestion, the irrelevant reward was used in order to rule out the possibility that fractional anticipatory responses rather than the drive stimuli themselves would mediate learning. While both end boxes contained goal animals in order to control for odor cues, the goal-box door on the incorrect side was locked. Two male goal animals were used and assigned randomly to the goal boxes each day. The C *S*s were allowed to retrace after an incorrect initial choice to the correct arm and receive the social reward for 30 sec. After the entry of the *S* into the goal box, the goal animal was released from behind the transparent partition. The NC *S*s were not allowed to retrace and received the reward only after correct initial responses. After incorrect choices the NC-A *S*s were retained in the incorrect arm of the alley for 30 sec., while the NC-M *S*s were retained in the mesh chamber for the same length of time. The door to the mesh chamber was not opened until the guillotine door was closed behind the *S*. The guillotine doors were used for all groups to prevent retracing after initially correct responses.

The NC-M group was included in the present experiment because it was of interest to determine the effect on learning of escape from the nonreward environment. However, some *S*s delayed several minutes before entering the mesh chamber, and in some cases this behavior persisted throughout the experiment.

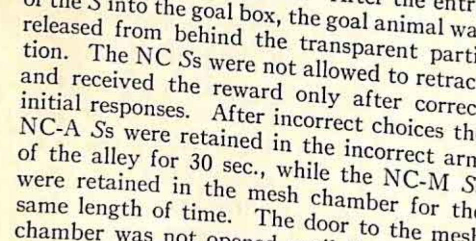


FIG. 2. Percent correct responses on first daily trials in 6-day blocks.

The decision was made to terminate the experiment on Experimental Treatment Day 60 or 84. If one or more groups of *S*s had been performing above 70% for 4 wk. or more at Day 60, the experiment would be terminated. If the criterion had not been met, the experiment would be continued until Day 84. This decision was partly based on Levine's (1953) finding that his *S*s had reached asymptote when they were averaging 75% correct responses.

RESULTS

The percentage of correct responses for all groups in successive blocks of six first trials is shown in Fig. 2. At least for the NC-A group, learning appears to be present in the last five 6-day blocks. Thus for group comparisons, the number of correct responses in the first five and the last five 6-day blocks for each *S* is presented in Table 1. For ease of comparison, *S*s are ranked in each group according to correct responses in the last five 6-day blocks. With the assumption that the probability of a chance success on a single trial is .5, the Bernoulli theorem was used to compute for each animal the probability associated with the number of correct responses in 30 trials. Only the number of correct responses which yielded probability values at or below the .05 and .01 levels are indicated in Table 1. While 9 of the *S*s in the NC-A group performed significantly above chance during the last five 6-day blocks, only 3 and 1 in the NC-M and C groups, respectively, gave any indication of learning. However, if the last four 6-day blocks is used for the analysis, significant p values are obtained for the performance of 4 of the NC-M *S*s while the results for the other groups remain the same. According to Wilkinson (1951) 4 or more significant ($p < .05$) statistics out of 10 possible are sig-

TABLE 1

Correct Responses for Each Animal on the First Daily Trials for Each Half of the Learning Period

No. Correct Responses					
Group C		Group NC-M		Group NC-A	
Days 1–30	Days 31–60	Days 1–30	Days 31–60	Days 1–30	Days 31–60
27**	20*	18	24**	16	24**
13	19	22**	20*	12	24**
16	17	17	20*	18	23**
14	17	17	19	19	22**
15	16	13	19	18	22**
14	16	19	18	14	22**
16	15	19	18	18	21*
15	15	17	18	16	21*
15	14	15	16	12	21*
16	12	11	14	16	18

$^{*}p < .05$.
$^{**}p < .01$.

nificant at the .001 level of confidence. The percentages of correct responses for the *S*s in the last five 6-day blocks were transformed by the arc-sine method and subjected to a simple analysis of variance. The main effect of groups was significant, F (2, 27) = 15.57, $p < .001$. According to the t tests, all three groups differed significantly ($p < .025$) from each other.

In order to examine possible tracking behavior, the number of correct responses on the first daily trials was arranged in the same order in which the *S*s followed each other in the experiment. From an inspection of the data, it did not appear that the number of correct responses was related to running order. Since a constant running order was used, it is not possible to adequately evaluate tracking behavior in the present experiment. However, if the number of same or different responses was the same for *S*s which were adjacent in the running order, the occurrence of tracking behavior is not likely. The first daily trials of Days 1–25 were used for the comparison since performance averaged 53% correct during this period. The average percentage of same responses during the period from Days 1–25 was 54%. Summing over trials, the *SD* for adjacent *S*s was 9.7% and summing over *S*s, the *SD* for trials was 10.4%. It may be concluded that, at least in the early portion of the experiment, tracking behavior was not exhibited.

Alternation from a correct response on Trial 1 to an incorrect response on Trial 2 of each day did occur in all three groups. The percentage of *S*s alternating from correct to incorrect increased from a chance level in the first week of the experiment to high levels in the following 5 wk. followed by chance levels in the last 4 wk. More specifically, the C, NC-A, and NC-M groups, respectively, alternated 72, 68, and 68% of the time during the first 6 wk. of the experiment and 48, 45, and 55% during the final 4 wk. Alternation from incorrect to correct responses was shown only in the NC groups; about 50% of the C *S*s responding incorrectly on Trial 1 of each day responded correctly on Trial 2. This 50% level was maintained fairly consistently over weeks; a small increase, from 48% to 62% was seen from the first 7 to the last 3 wk. The NC-A and NC-M groups, respectively, averaged 85% and 75% alternation from incorrect to correct, with little variation seen over weeks.

For the later trials of each day, alternation from correct to incorrect responses decreased substantially for the NC groups, while the C group continued to alternate as in Trials 1 and 2. The alternation from incorrect to correct responses increased for the NC groups in the later trials of each day and remained the same as in Trials 1 and 2 for the C group.

Figure 3 presents the percentages

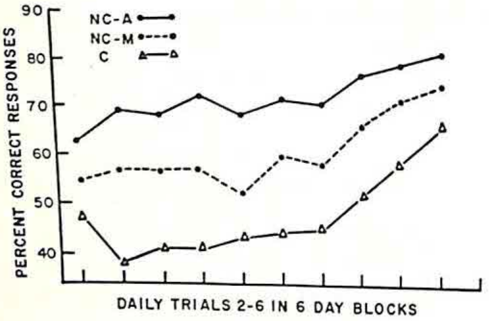


FIG. 3. Percent correct responses on Daily Trials 2–6 in 6-day blocks.

of correct responses for all groups on Daily Trials 2–6 in 6-day blocks. Performance levels on Trials 2–6 ranked for the groups in the same order as the Trial 1 data in the last half of the learning period. The alternation from correct to incorrect responses for the NC groups from Trial 1 to Trial 2 of each day was not sufficient to depress Trial 2–6 performance nor even Trial 2 performance below chance. In addition, the high percentages of alternation from incorrect to correct responses for these groups resulted in performance levels well above chance. For the C group, the high percentage of alternation in the early part of the experiment from correct to incorrect responses on all trials of each day coupled with the failure to alternate from incorrect to correct responses resulted in performance levels below chance for the early part of learning. It is also apparent that the transfer of *S*s to individual cages on Day 46 was accompanied by an increase in the percentage of correct responses for all groups. It will be recalled that the amount of alternation from correct to incorrect responses decreased for all groups after Day 46 on all trials of each day.

In order to determine if the same relationship held between Trials 1 and Trials 2–6 for the *S*s within each group as between groups, several comparisons were made. For each *S*, the total correct responses on the first daily trials in the last half of the learning period were compared with the total number of correct responses on Trials 2–6 of each day for the first half of the learning period and for the total learning period. It was clear from an inspection of the data for each group that no relationship was present.

The mean speeds on the correct and incorrect trials for each group were calculated for the first daily trials and Trials 2–6 separately. Since no differences were found between groups nor between the speeds on the correct and incorrect trials, these data are not presented. The general trend of the speeds was an increase up to 4 wk., followed by a decrease for 3 wk. and an increase in the last 3 wk. The increases in the speeds in the final weeks may have been related to the transfer of *S*s to individual cages.

DISCUSSION

A comparison between the C and NC-A groups is sufficient to evaluate the hypothesis; from this comparison the noncorrection method is clearly superior to the correction method in promoting learning on the basis of drive stimuli. As mentioned earlier, some of the NC-M *S*s did not enter the mesh chamber readily, making any interpretation of the results for the NC-M condition questionable.

Of particular interest is the comparative performance of the groups on Trials 2–6 of each day, which paralleled the rank order of group performance on the first daily trials. Thus the learning of the groups was directly related to the number of immediate reinforcements. It may be recalled that the learning of *S*s within groups did not show this same relationship. The number of reinforcements after initially correct responses may not be the only critical factor, however. Since the NC groups were never reinforced for incorrect responses while the C group was always reinforced, it seems likely that the approach tendencies to the incorrect side were being weakened for the NC groups and strengthened for the C group. If the relative strengths of the correct and incorrect response tendencies determine choice behavior, then the NC condition would be expected to be superior to the C condition.

Since for the C group the delay of reward was longer after incorrect responses than after correct responses, it would be expected that the *S*s would eventually choose the shorter route to the goal. However, even on Trials 2–6 of each day the C group consistently performed below chance prior to being transferred to individual cages. Perhaps the stimulus satiation effect suggested by Glanzer (1953) is relevant to this finding. While Glanzer in his discussion eliminated situations where a reward is used, the present situation may possibly be meaningfully evaluated from his point of view. While the NC groups did alternate consistently above 50% from incorrect to correct responses, the C group did not. It seems likely that since the C group was allowed to correct after incorrect responses and therefore went to the positive side last, the satiation effect for the correct side was thus greater or at least equal to that for the incorrect side. If we assume that the social drive was relatively weak, consequently leading to a negligible difference between the excitatory strengths for the two sides, then the relative amounts of satiation on the previous trial in some cases would result in the dominance of the incorrect over the correct tendency. Thus the 50% alternation from incorrect to correct responses obtained in early learning for the C group does not seem unreasonable. In addition, increasing the drive by transferring animals to individual cages would lead to an increased difference in the two excitatory strengths so that the satiation would be negligible in its effect on choice behavior. Therefore the obtained result in the later trials of increased alternation from incorrect to correct responses would be expected.

REFERENCES

GLANZER, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60**, 257–268.

HULL, C. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.

KALISH, D. The non-correction method and the delayed response problem of Blodgett and McCutchan. *J. comp. Psychol.*, 1946, **39**, 91–107.

LEEPER, R. The role of motivation in learning and a study of the phenomenon of differential control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.

LEVINE, S. The role of irrelevant drive stimuli in learning. *J. exp. Psychol.*, 1953, **45**, 410–416.

WEBB, W. B. Drive stimuli as cues. *Psychol. Rep.*, 1955, **1**, 287–298.

WILKINSON, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156–158.

(Received June 13, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 420–421

SUPPLEMENTARY REPORTS

COMPARISON OF TWO PROCEDURES IN THE STUDY OF RETROACTIVE INTERFERENCE IN CONNECTED MEANINGFUL MATERIAL[1]

DAVID J. KING AND SARAH TANENBAUM

American University

The Individual and Group conditions of experimentation in retroactive interference (RI) were contrasted. In the Individual condition, 3 groups (1 control, 2 experimental) of 12 Ss each were tested individually with a memory drum exposing the material and learning undertaken according to the method of serial anticipation. In the Group condition, similar groups of Ss were tested on the same learning material under group conditions with unlimited time available for written recalls. 2 analyses of covariance were significant at the .01 level. Relative amounts of RI were nearly identical for both sets of experimental groups. The present data support the conclusion that RI can be produced with connected meaningful material in group presentation and with unlimited recall time.

Slamecka (1959) has reported the production of significant amounts of retroactive interference in connected meaningful material using group testing procedures. In a later article (Slamecka & Ceraso, 1960) however, it was concluded that it has been difficult to produce significant amounts of retroactive interference in such material because most studies have employed group testing, whole presentation, unlimited time recall, and other such low control techniques of experimentation. These conclusions were primarily based on the results of two experiments (Slamecka, 1960a, 1960b) where connected meaningful material yielded to significant amounts of retroactive interference. The purpose of this experiment is to contrast the experimental techniques used by Slamecka in the above two studies with the more usual procedure utilizing, however, learning materials selected from those used by Slamecka.

Method.—The *S*s (university students) were divided into two major conditions and six groups, three groups each for the Individual condition and the Group condition. Each group contained 8 females and 4 males. The OL and IL material consisted of the second pair of sentences from Slamecka's (1960a) study. The first sentence was used for OL and the second for IL. All 72 *S*s received two OL trials. For both general conditions, groups were defined in terms of the IL procedure. Groups A and B, the four experimental groups, were given eight IL and four IL trials, respectively. The control groups C were given a period of color naming during the IL interval. This activity was also used to fill the temporal gap for the B groups. All three groups in the Individual condition were given four delayed recall trials, while in the Group condition only one written delayed recall was required. The Individual condition utilized memory drum presentation of the method of serial anticipation. The general running and scoring conditions were similar to those employed by Slamecka (1960a). In the Group condition all *S*s in a given group were run together with the stimuli read to the *S*s and written recalls (unlimited time) obtained. Time intervals were approximately the same as in the Individual condition. Scoring criteria were the same as employed by King and Cofer (1960). Recall measures (either anticipations or written recalls) were also obtained after presentation of the OL and IL (except control groups) material.

Results and discussion.—The data (percentages transformed to arc sines) for the two general conditions were subjected to analysis of covariance, using OL scores as the adjusting variable. The assumptions for analysis of covariance were tested and met (i.e., the data for both general conditions exhibited homogeneity of variance and homogeneity of within-groups regression). Both covariance *F* tests were significant at better than the .01 level—Individual condition, $F\ (2, 32) = 7.49$, $p < .01$; Group condition, $F\ (2, 32) = 8.51$, $p < .01$. Following the method proposed by King (1960) to express the magnitude of retroaction, the two sets of delayed recall means were adjusted with respect to both the original recall means and the within-groups regression coefficients. The results of these

---

[1] This research was submitted by the second author as a directed Master of Arts degree research project. The study was designed and supervised by the first author.

TABLE 1

MEAN VALUES FOR ORIGINAL LEARNING AND RECALL, AND PERCENT RETROACTIVE INTERFERENCE

Group	Mean		Percent RI
	Original Learning	Recall	
Individual condition			
A	38.12	20.56	43.76
B	35.78	24.33	27.65
C	35.14	33.32	0
Group condition			
A	48.53	20.57	40.15
B	49.87	23.16	31.47
C	44.58	33.14	0

adjustments gave the delayed recall means the magnitude that they would have had if the original recall means had been identical in magnitude. The adjusted delayed means for the control groups were then taken as zero retroaction and the adjusted delayed recall means of the experimental groups were expressed as a percentage of the control groups to indicate the magnitude of retroactive interference. The results of this analysis, together with the unadjusted original learning and recall means, are presented in Table 1. The *t* tests between means of the two experimental groups were not significant, but both experimental groups differed from the control at the .01 level of significance for a one-tailed test. Levels of significance were determined by using Fisher's adjusted probabilities (Edwards, 1956). Clearly, the same amount of interference was generated by both procedures.

Slamecka's suggestion that rigorous control procedures are needed to produce significant amounts of retroactive interference in connected meaningful material may be correct, but in view of the above results, and Slamecka's (1959) earlier work, the evidence to support such a conclusion seems inadequate.

REFERENCES

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1956.

KING, D. J. Expressing the magnitude of results in retroaction experiments. *Psychol. Rep.*, 1960, **7**, 312.

KING, D. J., & COFER, C. N. Retroactive interference in meaningful material as a function of the degree of contextual constraint in the original and interpolated learning. *J. gen. Psychol.*, 1960, **63**, 145–158.

SLAMECKA, N. J. Studies of retention of connected discourse. *Amer. J. Psychol.*, 1959, **72**, 409–416.

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of practice level. *J. exp. Psychol.*, 1960, **59**, 104–108. (a)

SLAMECKA, N. J. Retroactive inhibition of connected discourse as a function of similarity of topic. *J. exp. Psychol.*, 1960, **60**, 245–249. (b)

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

(Received June 13, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 4, 421–422

A REPLICATION OF FACILITATION OF CONCEPT FORMATION THROUGH MEDIATED GENERALIZATION[1]

JERRY HIGGINS, SARNOFF A. MEDNICK, AND SUSAN L. TAYLOR

University of Michigan

100 undergraduate Ss learned 1 of 2 paired-associate lists of 12 pairs and were then presented with a list of 12 nouns and were instructed to discover into what 3 groups the nouns could be divided and what adjective could describe each group. The paired-associate lists were designed by a method based on mediated generalization to facilitate a given concept in the concept formation list. All results were in the direction of the predicted facilitation; all but 1 were statistically significant beyond the .01 level. It was concluded that the replication was successful.

Mednick and Freedman (1960) have reported that the prior learning of a paired-associate list facilitates the attainment of selected concepts when the paired-associate list is constructed in accordance with a mediated generalization paradigm. Thus, for example, in order to facilitate the production of the concept response WHITE to the concept instance SUGAR prior pairing of the words SUGAR-BLACK was given in a paired-associate

---

[1] This research was facilitated by the Cooperative Research Branch of the Office of Education, United States Department of Health, Education, and Welfare through Contract 1073 to Sarnoff A. Mednick and Martha T. Mednick and by a predoctoral fellowship from the National Institute of Mental Health, United States Public Health Service to J. Higgins.

list. It was assumed that the word BLACK elicited the implicit response WHITE causing the word WHITE to occur in contiguity with the word SUGAR, strengthening the association SUGAR-WHITE. When *S* was subsequently presented with the instance SUGAR in the concept formation task, the attainment of the concept WHITE was facilitated.

The present study attempted to replicate this result with an *N* of 100 undergraduate students consisting of 43 males and 57 females.

Results.—The mean number of trials to attain the facilitated and the nonfacilitated concepts was 12.83 and 13.70, respectively. This difference was not significant, t (99) = 1.32, $p > .10$. The mean number of errors made on the facilitated and nonfacilitated concepts was 33.05 and 41.61, respectively, a significant difference, t (99) = 3.18, $p < .005$. The mean trial on which the first correct instance of the facilitated and nonfacilitated concepts occurred was 5.84 and 8.43, respectively, a significant difference, t (99) = 2.78, $p < .01$. In addition to the above analyses performed by Mednick and Freedman, an analysis was made to determine whether the instances of the facilitated concepts were more easily elicited during the early trials. The mean number of different correct instances for the facilitated and nonfacilitated concepts in the first four trials was 1.64 and 1.09, respectively, a significant difference, t (99) = 2.75, $p < .01$. Further, the number of *S*s showing more different facilitated than nonfacilitated instances in the first four trials, showing an equal number, and showing fewer was 49, 27, and 24, respectively, χ^2 (2) = 11.18, $p < .005$. In computing this chi square we conservatively assumed a chance distribution of *S*s in the three categories.

It may be noted that Mednick and Freedman's *S*s were somewhat superior in the ability to form both facilitated and nonfacilitated concepts to the *S*s used in the present study. This difference may be attributable, at least in part, to the nature of the two groups. The *S*s in the original study were unpaid volunteers, while the *S*s of the present study were paid volunteers and students fulfilling a requirement of an introductory psychology course.

In summary, all results of the replication experiment were in the direction of facilitation; all but one were statistically significant. It is concluded that the facilitation of concept formation through mediated generalization is a replicable phenomenon.

REFERENCE

MEDNICK, S. A., & FREEDMAN, J. L. Facilitation of concept formation through mediated generalization. *J. exp. Psychol.*, 1960, **60**, 278–283.

(Received June 18, 1962)

SCIENTIFIC PROTOTYPE OFFERS YOU A CHOICE OF THREE HIGH CAPABILITY TACHISTOSCOPES TO FILL ALL YOUR RESEARCH REQUIREMENTS

Scientific Prototype's line of two- and three-channel Tachistoscopes offers you the highest order of precision, range, and repeatability for your research.

The designs feature unrivaled timer accuracy of time interval generation from 0.1 millisecond up to 110 seconds, and unique lamps with 2½ microsecond maximum rise time, and up to 300 foot-lambert output. Many other features are incorporated that increase the capability and flexibility of the units.

The Scientific Prototype Tachistoscopes are ruggedly built. The logic circuitry is all transistorized, and up to 1000 fold "worst case" derating is applied to all components. Mil. Spec. or Computer Grade components are used extensively.

An extensive selection of sophisticated accessories is available to further increase the utility of these Tachistoscope designs.

Write for complete data on the Tachistoscope models and accessories available.

PARTIAL LIST OF SCIENTIFIC PROTOTYPE TACHISTOSCOPE USERS:

Creedmore State Hospital, Princeton University, Education Testing Service, Columbia University, United States Navy, Duke University, University of Arizona, National Aeronautics and Space Administration, Clarke University, Longview State Hospital, Massachusetts Institute of Technology, United States Air Force, International Business Machine Corp., Yale University, Johns Hopkins, Van Etten Hospital, Boston VA Hospital

SCIENTIFIC PROTOTYPE MFG. CORP.

Behavior Analysis Apparatus . . . operant conditioning kits, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, times, switches, digital test equipment and special systems.

623 W. 129th STREET, NEW YORK 27, NEW YORK
TELEPHONES (212) MO 6-4674 and 4789

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland
PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich
PROF. DR. H. ROHRACHER, Wien

Herausgeber für die Schweiz
PROF. DR. R. MEILI, Bern

Redaktion
DR. C. J. HOGREFE, Göttingen

Band IX/3

RAINER FUCHS, Schmerz als Signal für die Bedrohtheit von Leib und Leben

THEO HERRMANN, Syntaktische Untersuchungen zum unmittelbaren Behalten

GERHARD KAMINSKI UND UTE OSTERKAMP, Untersuchungen über die Topologie sozialer Handlungsfelder

WOLFGANG SCHÖNPFLUG UND MANFRED SCHÄFER, Retention und Aktivation bei akustischer Zusatzreizung

FRIEDRICH SIXTL, Ein Vergleich der Prozent-Quotient-Methode von Clostermann mit anderen Vier-Felder-Korrelationsmaßen

REINHARD TAUSCH, Merkmalsbeziehungen und psychologische Vorgänge in der Sprachkommunikation des Unterrichts

ERNST TIMAEUS, Über den Aufbau einer Lochkartenkartei psychologischer Literatur

Kritische Buchbesprechung:
JOHANNES LINSCHOTEN, Auf dem Wege zu einer phänomenologischen Psychologie. Die Psychologie von William James (W. D. Fröhlich)

All articles are provided with summaries in English and French.
The journal appears quarterly and contains 160 pages per number.

Subscription is DM 64,—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
(Brentanoweg 10, Germany)

Journal of
Experimental Psychology

Vol. 65, No. 5 May 1963

COGNITIVE ASPECTS OF INFORMATION PROCESSING: II. ADJUSTMENTS TO STIMULUS REDUNDANCY[1]

PAUL M. FITTS, JAMES R. PETERSON, AND GERSON WOLPE

University of Michigan

3 experiments are reported in which relative stimulus frequencies were varied in 9 choice tasks. The tasks involved naming numbers and pointing to lights. It was found that as redundancy increased average RTs to the frequent stimulus component decreased whereas RTs to less frequent components increased, the differences being a linear function of redundancy. These effects were greater for the less compatible (vocal) task. Ss used the frequent response more often and the infrequent response less often than appropriate in responding to redundant sequences. These results are in agreement with predictions from a stimulus sampling and sequential decision model in which it is assumed that RTs and errors are a function of prior probabilities and the payoff matrix for correct and wrong, slow and fast responses, as well as a function of stimulus discriminability.

A process that generates sequences of discrete events is said to be redundant when all possible sequences are not equally probable. Adjustments to the more proximal redundancies in a stimulus sequence is an important form of learning and can be observed in many forms of perceptual, motor, and verbal behavior. Thus many years ago J. McKeen Cattell (1885, 1886) noted that serial response rate in vocalizing the names of individual letters in a list is about twice as fast for familiar (forward) sequences taken from English text as for identical sequences presented in an unfamiliar (backward) direction. Comparing redundant sequences from different languages he (Cattell, 1947) concluded further that "the less familiar a language, the smaller the difference between the speeds of reading a selection forward and backward" (p. 24).

More recently, Hyman (1953) and Crossman (1953), working independently, have both found that average RT across all stimulus components is reduced in proportion to average stimulus redundancy, the size of the stimulus set remaining constant. With respect to individual response times, Hyman (1953) noted further that ". . . the reaction time to the low information component was mark-

---

[1] This research was supported by the United States Air Force through the Office of Scientific Research, under Contract No. A F49 (638)-449.

edly lower than would be predicted on the basis of the regression line fitted to the means of the conditions" and concluded that ". . . if . . . we are interested in the behavior of the components making up the conditions, we must find different laws and equations" (p. 194) than those used in predicting overall mean response times.

For the most part the models that have been proposed to account for time and error data in choice RT experiments have neglected the problem of the nature and bases for adjustments to stimulus redundancy. These models are of two classes, (*a*) search models and (*b*) stimulus sampling models. In the first, the assumption is made that incoming signals are identified by comparison with stored replicas, these comparisons requiring more time, on the average, the more extensive the search process required to locate the stored information; these models either assume "errorless" performance or otherwise neglect time-error relations. In the second, the assumption is made that *S* samples the stimulus, which is always accompanied by "noise," identification being accomplished by a statistical decision process. The latter model predicts that more similar stimuli, smaller signal-to-noise ratios, and fewer errors will be associated with longer mean RTs, a prediction which is consistent with the finding that increased time can be used to achieve greater accuracy in a variety of perceptual-motor tasks (Egan, Greenberg, & Schulman, 1961; Fitts, 1954; Swets & Green, 1961). There may be nothing logically inconsistent, however, in the assumption that RT is the consequence of a two-stage process, which involves time for access to stored information followed by statistical decision time.

Welford (1960), considering redundancy in the context of a general discussion of search models, concluded that a process of "serial dichotomous classification with check on accuracy" should result in faster responses to the more frequent events in a redundant sequence, but that RT for a less likely event in the same sequence should never be more than twice as long as that for a more likely event. Stone (1960) has proposed a sequential sampling model for choice RT, basing it in part on Wald's (1947) theoretical treatment of sequential decision making and on the general theory of signal detection in noise (Swets, Tanner, & Birdsall, 1961), and in part on empirical data regarding the effects of discriminability on RT (Crossman, 1955; Crossman & Szafran, 1956). Stone concludes that his model can be extended to the prediction of RTs to redundant sequences, and makes the important point that as error rate becomes small, the cost in time required to achieve additional reductions in errors should become increasingly great.

The sequential sampling model appears to offer some promising leads for the prediction of adjustments to redundant stimulus sequences. A brief summary of the major assumptions and some qualitative predictions from this model follow.[2]

Assumptions regarding a probability ratio computer and decision maker.—A process is assumed which operates in a manner analogous to an efficient probability computer and decision maker which bases its decision on posterior probabilities computed in accordance with Bayes' theorem (see Edwards, Lindman, & Savage, 1963; Raiffa & Schlaifer, 1961). Such a computer is assumed to know the initial odds governing the occurrence of alternative stimuli, as well as the conditional distribution of an input parameter, X, which represents the data available from independent fixed-length samples of the stimulus. New (posterior) odds are computed following each successive observation. As a consequence of sampling variations and internal (biological) noise the sequence of values computed in this manner should have the properties of a random walk, but should drift in a direction determined by the characteristics of the particular stimulus being observed. It is further assumed that this sequential observation process will continue until the odds reach some cutoff point such that the expected value of terminating the process (and making a response) exceeds the expected value of continuing to observe.

---

[2] Appreciation is expressed to Ward Edwards and Wilson P. Tanner, Department of Psychology, and Wilfred Kincaid, Department of Mathematics, University of Michigan, for helpful criticisms and suggestions regarding this section.

Given symmetrical cutoff points (such as would result from the use of equal weights for all types of errors) but unequal prior probabilities, the random walk would begin nearer one cutoff point than another and should therefore terminate sooner, on the average, when the probable stimulus occurs than when the less probable one occurs. This would result in faster average RTs to the more probable stimulus.

In summary, RTs to different stimulus components should be a function of the initial stimulus probabilities, the intrinsic characteristics of the successive observations, and the cutoff points; whereas error rates should be determined solely by the cutoff points. The cutoff points, in turn, are a function of the payoff matrix that specifies (*a*) the costs and values, respectively, of each kind of error and each kind of success, and (*b*) corresponding costs and values for fast vs. slow responses.

It is possible to base the strategy of such a decision process on error rates computed in one of two ways. Consider, as an illustration, two alternative stimuli, S_a and S_b, and two responses, R_a and R_b. From the viewpoint of conventional hypothesis testing one usually computes the probability of an error, given that S_b occurred, $P(R_a|S_b)$, and the probability of an error given that S_a occurred, $P(R_b|S_a)$. As an example, assuming the hypothetical frequencies shown in Table 1, where S_a occurs nine times as often as S_b, these two error rates would be equal and would equal the overall error rate (.10). The alternative is to estimate the probability that R_a will be in error when it occurs, considering all previous instances of this response, $P(S_b|R_a)$, and the probability that R_b will be in error, $P(S_a|R_b)$. In the example given in Table 1, the error rates computed in this second way are very different indeed, .0122 and .50, respectively.

It is assumed that after considerable experience with the error rates in a given task, *S*s may adopt a strategy which can be predicted from odds computed by the second of the above procedures. This is equivalent to *S* asking, "If I make Response A what are the chances of being wrong?" Given unequal stimulus probabilities such as those shown in Table 1, and instructions to minimize overall errors and RT, this assumption suggests a strategy whereby the frequency of A responses would be increased, and B responses decreased relative to the frequencies shown in Table 1. This could be accomplished by appropriate shifts either in the cutoff points or

TABLE 1

Hypothetical Frequency Table

Response	Stimulus		Subtotal
	S_a	S_b	
R_a	81	1	82
R_b	9	9	18
Subtotal	90	10	

in the initial odds point. Such a strategy should reduce average RTs even if the overall error rate was unchanged, since (A) responses would now occur even more often and would become even faster.

These ideas are essentially a reformulation of some theories regarding expected value maximization proposed initially by such historical figures as Bernoulli, Bayes, and Laplace. Stated in terms of expectancy theory the more strongly *S* expects Event A, the wider will be the range of events he is likely to accept as signaling the occurrence of A and the faster his A responses, whereas the more likely he is to fail to respond correctly or rapidly to contrary evidence.

Without attempting a formal demonstration of the predictions from a sequential stimulus sampling model at this time the hypothesized effects of increased redundancy on speed and the accuracy of choice responses, values, and costs, as well as the nature of the stimulus set remaining fixed, can be summarized as follows:

Reaction time.—It is conjectured that as the prior probability of one stimulus in a sequence is increased relative to alternative stimuli the time spent observing prior to reaching a decision, when a frequent or expected stimulus occurs, should decrease, and the time spent observing when a less frequent stimulus occurs should increase, on the average, as a function of redundancy.

Errors.—It is conjectured that as the prior probabilities with respect to alternative stimuli are made increasingly different, *S*s will use the more probable response more often and the less probable responses less often than the corresponding stimulus frequencies, but that the relative frequencies of errors, computed in terms of response frequencies, will remain essentially constant for all responses. However, when the prior probability favoring one of the possible stimuli is made sufficiently large, *S*s should cease to attempt to discriminate and make only the response appropriate to that stimulus.

Three experiments were conducted to test these predictions. These are part of a program of research on cognitive factors in information processing (see Fitts & Switzer, 1962), in which human information handling is assumed to involve processes analogous to the operations specified by a computer program, such as the use of subroutines based on a knowledge of the probabilities governing events in a particular task or situation.

Method

Three experiments,[3] involving vocal (naming) responses to numerals, and motor (pointing) responses to lights, were carried out. Number of alternative stimuli and responses was held constant at 9 in all three experiments, while redundancy was varied by increasing the relative frequency of one of the stimuli in the set while reducing the relative frequencies of the remaining eight stimuli. Nine rather than a smaller number of alternatives was used because it was felt that this would make it more feasible to collect data where very low probabilities were assigned to individual stimuli.

Experiment I

Apparatus.—Stimuli were $\frac{1}{4}$-in. high black numerals on white cards. They appeared in

---

[3] Gerson Wolpe, Gail Switzer, and Betty Brown Ames collected the data in Exp. I and II; James Peterson was responsible for data collection in Exp. III.

TABLE 2

LEVELS OF REDUNDANCY USED IN EXP. I

Cond.	Frequencies per Run of 126 Events		Relative Redundancy
	Numeral 1	Numerals 2–9	
A	14	14 each	.00
B	30	12 each	.03
C	94	4 each	.50
D	118	1 each	.83

a 2 × 2 in. window at eye level in the center of a gray screen, and were exposed by the opening of a double-bladed electromechanical shutter immediately behind the window. The *S*s were asked to wear sound attenuating headsets and listen to broad band noise as a means of providing a uniform auditory environment. Vocal reaction time was measured by the use of a throat microphone, an electronic voice key, and a .01-sec. clock.

Subjects.—The *S*s were 48 male college students, paid by the hour.

Levels of redundancy.—A total of 126 stimuli were seen by each *S* during each session. Four levels of redundancy were used (see Table 2). Stimulus sequences for any given *S* on any given session were determined by a random draw without replacement. For Cond. A, B, and C separate draws were made for each run of 126 stimuli. In the D condition (83% redundancy), the draw was made over 378 events, the sampling procedure being such that each of the different infrequent stimuli occurred a total of three times over three sessions and there was a total of 8 infrequent stimuli in each session run of 126 events, but the 8 different infrequent events did not necessarily each occur once per session (this was done to prevent *S*s from predicting the last infrequent stimulus by elimination).

Procedure.—Each *S* was tested individually for four sessions, approximately 1 hr. in length. On Session 1 all *S*s performed under Cond. A (equal frequencies). Successive sets of 12 *S*s were then divided into three equal strata (high, medium, and low) on the basis of their RTs in responding to the numeral 1. The 4 *S*s within each stratum were then assigned randomly, one to each condition, in order to control for ability levels across groups. This procedure was repeated for successive sets of 12 *S*s until all 48 *S*s had been assigned to groups.

The *S*s assigned to Group A continued as before on subsequent sessions. Other *S*s were told that the numeral 1 would now occur more frequently than the other numerals, and were shown a bar graph which depicted the relative probabilities of each numeral in Cond. B, C, or D, whichever, was appropriate. Conditions for each of the four groups remained unchanged during Sessions 2, 3, and 4.

A "ready" signal was given 2 sec. before each stimulus exposure. The *E* told *S* his reaction time after each response. The interstimulus interval was about 10 sec.

Experiment II

The purpose of Exp. II was to test the generality of the results when 5 rather than 1 was the most frequent stimulus under the redundancy level of Cond. D. Otherwise the apparatus and procedure of Exp. II was identical to that of Exp. I. Twelve new *S*s, drawn from the same college population, were used. They will be referred to as Group E.

Experiment III

The purpose of Exp. III was to test the generality of the results for a very different set of stimuli and responses and under three levels of redundancy, with *S*s serving as their own controls.

Subjects.—The *S*s were 15 male college students who were paid for their time.

Apparatus.—Stimuli were nine neon lights, which were visible through $\frac{5}{16}$-in. holes in the center of $\frac{3}{4}$-in. metal rings. The rings, which served as targets for motor (pointing) responses, were arranged in a half circle, equidistant from a starting button. Electronic circuits were used to measure reaction time in breaking contact between the forefinger and the starting place, and movement time between the starting plate and the target rings. This type of switching circuit operates by means of a small current passed through *S*'s body, the level being below the threshold for feeling. Two .01-sec. timers were used.

Preliminary work indicated that the positions of the upper and lower arms, the visual fixation point, and the grouping of the targets all affected component reaction and movement times to the nine stimulus points. In the arrangement finally selected (Fig. 1), the upper arm was held comfortably at the side of the body, the forearm was horizontal and rotated 15° toward the median plane, and a visual fixation point was located between the center target light and the home button.

Levels of redundancy.—Three levels of redundancy, 0, 34, and 71%, were achieved

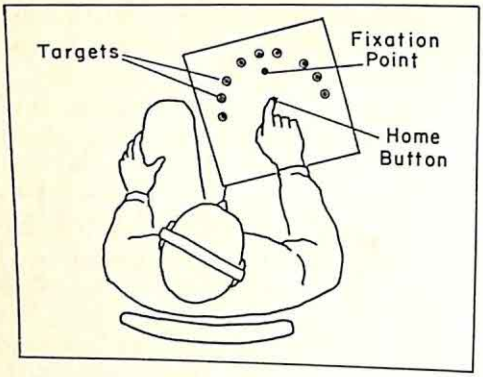


FIG. 1. Plan view of the spatial stimulus-response apparatus used in Exp. III.

by using stimulus sequences of a fixed length, and arranging that (*a*) the center and all other lights occurred 7 times each (in runs of 63), (*b*) the center light occurred 40 times and the remaining lights 24 times overall (in runs of 64), or (*c*) the center light occurred 56 times and the remaining lights 8 times overall (in runs of 64). In the *b* and *c* conditions the eight infrequent lights occurred equally often on the average over all sessions, but not necessarily equally often on each session; thus *S* was not able to predict the last few lights by elimination. Within the above restrictions the order of occurrence of stimuli was random and different for each *S* on each session. No restriction was placed on the length of runs of the same stimuli.

Procedure.—Each *S* served once under each of the three levels of redundancy during each session, making 191 responses per session, for a total of four sessions. The probability rules were explained, and a bar graph depicting the appropriate relative frequencies of different stimuli was visible during each run. The order in which each *S* worked at the three conditions on a given session was determined by the use of five different randomized 3 × 3 Latin squares. Knowledge of results was given at the end of blocks of 16 responses, and consisted of the times for the fastest and slowest responses in that block. A 2-min. break was given between conditions. Stimuli were presented at intervals of about 10 sec., 2 sec. after a ready signal. The *Ss* wore headsets and listened to masking noise.

RESULTS

Reaction times.—Choice reaction times (RTs) by sessions, pooled across all *Ss* in a group, but shown separately for more- vs. less-frequent components of the stimulus sequence, are summarized in Fig. 2 (Exp. I and II) and 3 (Exp. III).

In Exp. I a highly reliable difference was found between the vocal reaction time to the numeral 1 as compared to the other eight numerals even when all stimuli were equally frequent (Control Group A), the response to 1 being .027 sec. faster on the average. This difference, which is not present in simple RT data for numerals, has been found in another experiment (see Fitts & Switzer, 1962). The control group (Group A—equal frequencies) showed only a slight improvement in RT (about .02 sec. decrease) during the four sessions, the difference between 1's and the other eight numerals

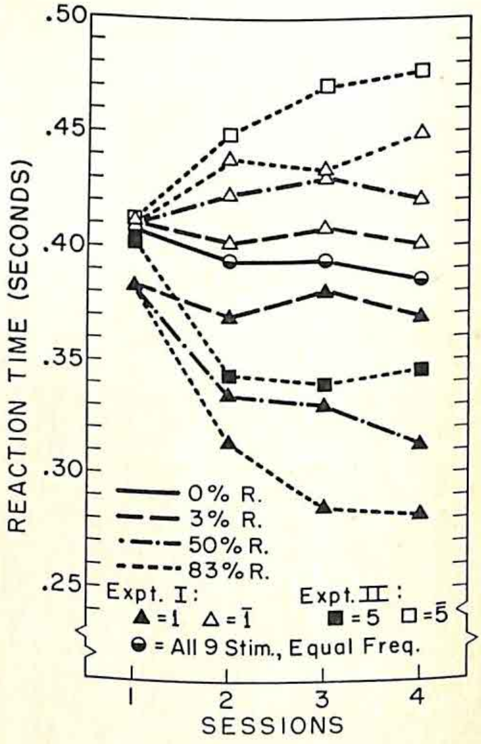


FIG. 2. Effects of levels of redundancy on RTs to stimulus components in Exp. I and II. (All groups received the equal frequency—zero redundant—sequence on Session 1.)

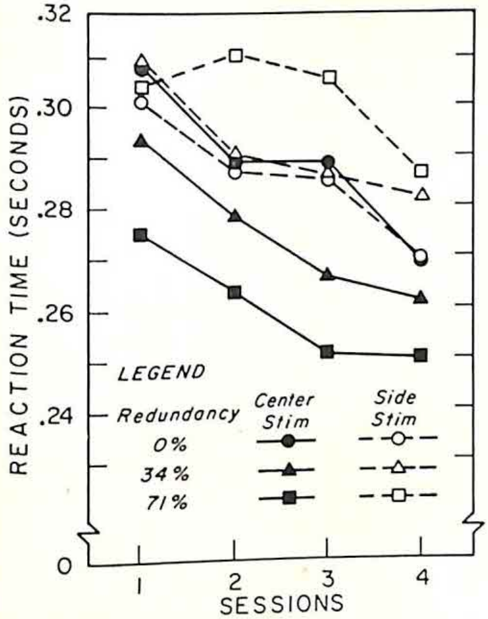


FIG. 3. Effects of levels of redundancy on RTs to stimulus components in Exp. III. (Each *S* worked at all three conditions on all 4 days.)

($\bar{1}$'s), remaining constant from session to session.

The major adjustment to stimulus redundancy occurred during the session in which redundancy was first introduced (Session 2), the magnitude of the effect being progressively greater the greater the degree of redundancy. Reaction times to the most frequent stimulus component became faster and RTs to the less frequent components became slower than under equal frequency conditions. This difference in RT to stimulus components tended to increase in magnitude as training under redundant conditions was continued. However, overall average RT (across all component stimuli) became faster with continued practice, and was somewhat faster on the average for the more redundant conditions.

The data for Exp. II, where the numeral 5 was the most frequent stimulus, are strictly comparable to those of Exp. I for similar levels of redundancy. For this group the RT for 5, under the equal-frequency condition of Session 1, was very close to the mean for the other eight numerals, as had been true for Group A; on subsequent sessions, after redundancy was introduced, the difference in the RTs to $\underline{5}$'s and $\bar{5}$'s for Group E was very similar in magnitude to the difference for 1's vs. $\bar{1}$'s of Group D in Exp. I after corrections are made for the initially shorter RT to 1.

Average RTs for the manual (pointing) responses in Exp. III (see Fig. 3), were much faster than had been found for the vocal responses in the preceding experiments and showed correspondingly smaller effects of redundancy. Under the equal frequency condition, the RTs for each of the nine lights were very similar, and decreased from a level of about .305 sec. to a level around .270 sec. with practice. As was previously the case for the most redundant sequences, RTs for the most frequent stimulus element (71% redundant) became progressively faster, and RTs for the less frequent elements became progressively slower, the maximum difference occurring on Session 2; during the last two sessions, responses to all stimuli, both frequent and infrequent, tended to become faster.

The effects of increased redundancy on the difference in RTs to frequent vs. infrequent components of a given stimulus sequence, are shown in Fig. 4, where data for the last 2 days have been averaged for each of the conditions of each of the three experiments. Each data point for vocal responses is for a different group of *S*s, except that the two data points for zero redundancy (1 vs. $\bar{1}$ and 5 vs. $\bar{5}$) are computed from the same control group (Group A). The three data points for

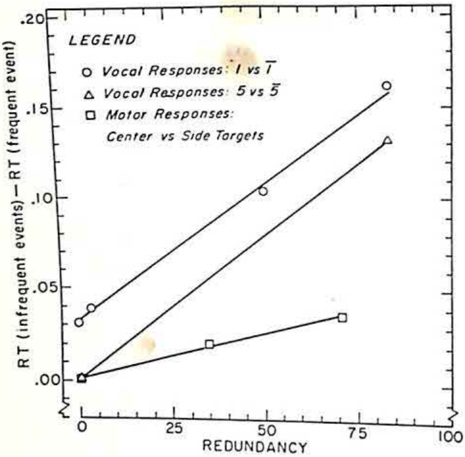


FIG. 4. Differences in RTs to infrequent vs. frequent stimuli as a function of the relative redundancy of the sequence.

Exp. III are for the same 15 *S*s serving under all conditions. It will be seen that the data for the first and third experiments are fitted very closely by straight lines, and the line connecting the two points for Exp. II has the same slope as for Exp. I. The intercepts reflect differences in RT for different stimulus components in the equal-frequency condition. The slopes for vocal response conditions are much steeper than for the manual (pointing) responses. In no case is the difference in RTs as great as the RT to the more frequent component, i.e., the ratio of RTs is less than 1:2, even for the most redundant condition.

Errors.—The errors made by *S*s in the last experimental session of Exp. I and II are summarized by stimuli in Table 2. The proportions of responses in all three experiments in which an inappropriate vocal response was made or the wrong target was touched were very small, just under 1%, and the overall error rate was relatively constant from session to session, improved performance showing up almost entirely as faster RT.

The 71 errors summarized in Table 3 include 17 instances in which an unintelligible sound, such as "ahhh" was uttered. These intrusions appeared to be randomly distributed among the conditions. Under Cond. A, where all stimuli occurred equally often, there were only 3 instances in which a wrong numeral name was spoken, and each involved a different confusion pattern. In terms of the stimulus sampling model discussed earlier the most pertinent error data are those for the highest redundancy level studied (Cond. D and E). Here there were 32 errors out of 3,017

TABLE 3

ERROR FREQUENCIES FOR THE DIFFERENT REDUNDANCY LEVELS OF EXP. I AND II ON THE FINAL EXPERIMENTAL SESSION

Cond.	Stimulus									Total Errors	Percent Errors
	1	2	3	4	5	6	7	8	9		
Exp. I											
A (R = .00)	0	1	1	2	1	1	0	0	5	11	0.7
B (R = .03)	0	3	1	0	2	0	1	0	0	7	0.5
C (R = .50)	2	3	1	2	2	1	3	0	0	14	0.9
D (R = .83)	2	5	0	1	4	2	2	1	2	19	1.2
Exp. II											
E (R = .83)	4	2	2	3	2	2	2	2	1	20	1.3
Total errors	8	14	5	8	11	6	8	3	8	71	0.9

Note.—The numeral 1 was most frequent in Cond. B, C, and D; 5 was most frequent in Cond. E.

responses (excluding unintelligible sounds) or almost exactly 1%. Of these errors, 31 were instances of using the most frequent response (saying "1" or "5") when an infrequent stimulus occurred, one was an instance of making an infrequent (wrong) response to an infrequent stimulus, and there were no instances of an error in which an infrequent response was given to a frequent stimulus. Stated in another way, 189 infrequent stimuli were presented by *E* (again excluding those to which an unintelligible response was made), but *S*s made only 158 infrequent responses. If we compute errors relative to stimuli, there were no errors (except "ahhhs") in responding to 2,985 frequent stimuli, but 32 or 17% errors in responding to infrequent stimuli. Thus error rates, computed in the conventional manner, were very unequal. Computed in the alternative manner, however, relative to responses, error rates were very similar. Thus the frequent response was used 2,859 times with 31 or 1.08% errors, and infrequent responses were used 158 times with 1 error (saying "6" instead of "7") or with a .6% error rate.

The two most redundant sequences had slightly higher average overall error rates than did the three less redundant series (1.25% vs. .70%) but the low incidence of errors makes any such comparison somewhat tenuous and the overall distribution of errors in Table 2 is not significantly different from chance. The point to be emphasized is that approximately the same proportions of each class of responses were in error, in spite of large differences in the probabilities governing individual stimulus events.

The overall rate of errors in Exp. III was almost exactly 1% and did not differ significantly among the three redundancy levels. Almost all errors, regardless of level of stimulus redundancy, involved lack of precision of movement rather than confusion among targets, i.e., errors involved touching an area near the correct target (most frequently to the left); instances of a clean hit on a wrong target were very few indeed (only about .2%).

DISCUSSION

The finding that differences in RTs are linearly related to redundancy in all three experiments is consistent with predictions from the present model. The error data also confirm the predictions, but the low error rates necessitate caution in making any generalizations about this aspect of the predictions. It appears that the model, which was discussed in terms of two-alternative choices, may be appropriate for situations where additional alternatives are present.

The finding of a smaller slope of the function relating redundancy to RT differences for the spatial vs. the symbolic S-R combinations agrees with the findings from several recent studies (Brainard, Irby, Fitts, & Alluisi, 1962; Crossman, 1956; Girew, 1958; Leonard, 1959) in which it has been found that variations in stimulus uncertainty have increasingly less effect on RT, the higher the initial S-R compatibility characterizing the task. The spatial S-R ensemble used in Exp. III is one of the most highly compatible that has yet been studied, since the mean RT to a nine-alternative, zero-redundant stimulus was only .27 sec.

It should be noted that many of the phenomena of "simple" RTs, such as the increase in RT resulting from the use of more "false" trials, the increase in RT resulting from an increase in foreperiod variability, and the decrease in RT and concurrent increase in anticipation error resulting from a shift from perceptua to motor set, are predicted by the presen model. In other words, the decision tha a no-stimulus situation has changed to stimulus situation is basically no differen than the decision as to which of the two stimuli has occurred. The present model is primarily a stimulus discrimination model, based on the assumption of the use of observation time and/or information processing time to increase response reliability. Perhaps the most critical conceptual issue is whether these stimulus sampling assumptions hold for stimuli that are ordinarily considered to be far above threshold.

The present experiments leave unanswered many questions concerning the processes involved in adapting to redundant sequences but in general indicate the validity of predictions about response times and errors derived from the stimulus sampling and sequential decision model. The sampling and decision processes assumed by this model are admittedly complex, but there is no reason to believe that the cognitive processes subsuming human behavior, even in the case of relatively rapid choice reactions, are necessarily any simpler.

REFERENCES

BRAINARD, R. W., IRBY, T. S., FITTS, P. M., & ALLUISI, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 321–329.

CATTELL, J. MCK. The inertia of the eye and brain. *Brain*, 1885, **8**, 295–312.

CATTELL, J. MCK. The time taken up by cerebral operations. *Mind*, 1886, **11**, 220–242.

CATTELL, J. MCK. *Psychological research.* Lancaster, Pa.: Science Press, 1947.

CROSSMAN, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

CROSSMAN, E. R. F. W. The measurement of discriminability. *Quart J. exp. Psychol.*, 1955, **7**, 176–195.

CROSSMAN, E. R. F. W., & SZAFRAN, J. Changes with age in the speed of information intake and discrimination. *Experientia*, 1956, Supple. No. 4, 128–135.

EDWARDS, W., LINDMAN, L., & SAVAGE, L. J. Beyesian statistical inference for psychological research. *Psychol. Rev.*, 1963, **70**, in press.

EGAN, J. P., GREENBERG, G. Z., & SCHULMAN, A. I. Operating characteristics, signal detectability, and the method of free response. *J. Acoust. Soc. Amer.*, 1961, **33**, 993–1007.

FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. exp. Psychol.*, 1954, **47**, 381–391.

FITTS, P. M., & SWITZER, G. Cognitive aspects of information processing: I. The familiarity of S-R sets and subsets. *J. exp. Psychol.*, 1962, **63**, 321–329.

GIREW, S. Information gain in tasks involving different stimulus-response relationships. *Nature*, 1958, **182**, 1819.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

LEONARD, J. A. Tactual choice reactions. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

RAIFFA, H., & SCHLAIFER, R. *Applied statistical decision theory.* Boston: Harvard Business School, Division of Research, 1961.

STONE, M. Models for choice-reaction time. *Psychometrika*, 1960, **25**, 251–260.

SWETS, J. A., & GREEN, D. M. Sequential observations by human observers of signals in noise. In Colin Cherry (Ed.), *Fourth London symposium on information theory.* Washington, D. C.: Butterworth, 1961.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychol. Rev.*, 1961, **68**, 301–340.

WALD, A. *Sequential analysis.* New York: Wiley, 1947.

WELFORD, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–229.

(Received July 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 433–443

SHORT-TERM RETENTION OF SINGLE PAIRED ASSOCIATES[1]

BENNET B. MURDOCK, JR.

University of Vermont

6 experiments investigated the short-term retention of single paired associates. Ss learned individual A-B pairs during 1 presentation of a list, then tried to recall the B member of the critical pair given A as the cue for recall. The first 2 experiments were replications of a previous study with number of prior pairs the parameter, the 2nd 2 experiments controlled for list length by using 6-pair lists, and the 3rd set of experiments attempted to assess the role of proactive inhibition (PI). Marked forgetting of single pairs consistently occurred within a retention interval of 6–10 sec., and increasing the number of prior pairs slightly decreased probability of recall. Very little interlist PI occurred, and the intralist PI decreased over time in contrast with the increase in PI over time that is usually found in the study of long-term retention.

A recent study of the retention of individual items (Murdock, 1961a) showed that recall probability dropped rapidly during a retention interval measured in seconds. The stimulus materials were either CCC trigrams or common English words, and the interpolated activity was either counting backwards or listening to a list of words of unknown length with instructions to recall the last three words. Although the forgetting was rapid, the results did not appear to be completely consistent with either a trace theory or an interference theory of forgetting.

The present research is a similar analysis of the short-term retention of single paired associates. The method used in this study has already been described in a brief note (Murdock, 1961b). Basically, a list of A-B pairs was presented once, then *S* was immediately tested for recall of one of the pairs with A as the cue for recall of B. With a fixed presentation rate (2 sec/pair), the number of pairs following the critical pair determined the length of the retention interval.

There seems to be little doubt that, with this procedure, most pairs are learned after a single presentation. That is, before an A-B pair is presented the probability that A as the cue for "recall" would evoke B is very low, since the words were paired at random. Immediately after the paired presentation recall probability is about .90 (Murdock, 1961b). Therefore, the decreased recall probability for delayed retention tests would indicate that forgetting had occurred.

During the retention interval *S*s are learning additional A-B pairs. Thus, one cannot distinguish between the effects of interference generated by these additional pairs and the effects of elapsed time per se. However, this procedure has the advantage of making rehearsal difficult because *S*s do not know what to rehearse (i.e., which pair they will be tested on).

Six experiments will be reported here. The first two are essentially replications of the previous study (Murdock, 1961b). The second two experiments are an attempt to study the same problem with a different experimental design. The third set of two experiments is an attempt to assess the possibility that massive PI accumulates when the same *S*s are repeatedly tested at all retention intervals.

---

[1] This work was supported by a research grant, M-3330, from the National Institutes of Health and by a training grant, G-12053, from the National Science Foundation. The author is particularly indebted to George F. Smith for analyzing some of the data, and to Jeanette Turner, Gretchen Miller, and Mary Jane Edwards for tabulating the results.

EXPERIMENTS I AND II

Method

Experiment I.—In general, the procedure here was identical with that used previously (Murdock, 1961b). A list of A-B pairs was presented once, then the A member of the critical pair was presented as the cue for recall of the corresponding B member. Naturally, the critical pair was in no way distinguished before or during list presentation. The retention intervals were 0, 2, 4, 6, or 10 sec., and at each retention interval there were either zero, one, two, or three prior pairs.

With this 5 × 4 orthogonal design one replication required 20 lists. In all, there were 16 replications (320 lists). All *S*s were tested on all 320 lists, 80 lists per session on each of four sessions.

The A and B members of the individual paired associates were English words, randomly selected from approximately the 4,000 most common English words (Thorndike-Lorge, 1944, G count of 20 or above). The words were randomly paired and randomly assigned to lists. With this sampling of stimulus material the results can be generalized to a known population of words.

All word pairs were written on individual index cards, photographed, and presented using a 35-mm. strip-film projector. The individual pairs were presented at a rate of one pair every 2 sec. The A member of the critical pair simply followed in sequence and without warning the last pair in the list, and it served as the cue for recall of B. To illustrate, the first list on Session 1 was: ROBE-SEVENTEEN; ROME-ASHAMED; EDUCATION-FIT; OFFER-KNOT; ROME? Thus, ROME-ASHAMED was the critical pair, the retention interval was 4 sec., and there was one prior pair. The correct response (i.e., ASHAMED) was given by only 2 of 20 *S*s.

Group testing was used, and *S*s wrote down their responses on numbered answer sheets. The *S*s were merely instructed to study each pair as it appeared and to do their best to recall the correct B word when the A word of the critical pair appeared on the screen. Two practice lists were given the first day; the first had a 4-sec. retention interval with one prior pair while the second had a 6-sec. retention interval with three prior pairs. No record of *S*'s recall on these two practice tests was obtained.

After each list there was a recall interval of 15 sec. This interval was terminated when *E* called out the number of the next list, and a verbal "Ready" signal preceded the start of the next list by about 2 sec. In all, approximately 20 sec. elapsed between the last pair in one list and the first pair in the next list.

The order of presentation of the 20 experimental conditions (five retention intervals and four levels of number of prior pairs) was randomized within each replication. With group testing the individual word pairs could not be counterbalanced among conditions. However, with random sampling of stimulus material and 16 replications of 20 different experimental conditions it seemed unlikely that systematic differences in item difficulty could bias the results.

There were 20 *S*s, and they were students of both sexes who were fulfilling a requirement of the introductory psychology course.

Experiment II.—In this experiment the retention intervals were 0, 2, 4, or 6 sec., but there were either zero, one, two, four, or eight prior pairs. Thus, the number of prior pairs was increased from a maximum of three in Exp. I to a maximum of eight in Exp. II.

With this 4 × 5 orthogonal design one replication again required 20 lists. In all there were 12 replications, 60 lists on each of 4 days. Group testing was used, and each *S* was tested on all 240 lists. There were 18 *S*s, all from the same group used in Exp. I.

New lists were constructed for this experiment. First, all homonyms, contractions, and archaic words were deleted from the original 4,000 words. Then, the words were randomly sampled and randomly paired; the only restriction was that there could be no more than 18 letters in any one pair. The pairs were typed in capital letters and then photographed, eight pairs to a frame. Seven-eighths of the strip-film projector window was masked so only one pair would appear on the screen at a time; the result was a rather effective "memory drum" for group testing.

For greater flexibility in subsequent experiments, the A member of the critical pair was not photographed; the strip film merely consisted of an unvarying succession of A-B pairs. In testing, when the end of a list was reached *E* blacked out the screen by closing a shutter placed over the projector lens. Also, *E* provided the cue for recall by calling out the A member of the critical pair.

There was one practice list on Session 1; the retention interval was 6 sec. and there were two prior pairs. Although told it was a practice list, *S*s were asked to write down their recall. There were 2 out of 18 *S*s who wrote down the correct response. The first test list was: HALT-LILY; STREAM-ASH; THINK-DEPART; STREAM? Thus, there was a retention interval of 2 sec., one prior pair, and 13 of 18 *S*s correctly recalled ASH.

In all other respects, the procedures of Exp. I and Exp. II were the same.

Results

The main results of Exp. I and II are shown in Table 1. As can be seen, the results of the two different experments were very similar. Also, the values are in rather close agreement with those previously reported (Murdock, 1961b, Fig. 1, p. 280). Thus, the retention curves appear to be reasonably reproducible with different *S*s and with different lists.

Marked forgetting occurred within a retention interval of 6–10 sec. Analyses of variance were performed on the number of correct recalls in Exp. I and Exp. II, and the results of these analyses are presented in Table 2. Retention interval (t) was highly significant in both experiments. Also, number of prior pairs (PP) was highly significant in both experiments, and the PP × t interaction was significant at the .01 level in Exp. I and at the .001 in Exp. II. The nature of the interaction can be best illustrated by reference to Table 1. The effect of retention interval was less with zero prior pairs than with one or more prior pairs. Also, with one or more prior pairs the retention curves were not perfectly regular, and this fluctuation would contribute to the interaction.

One method of analyzing retention data is to determine retention loss by comparing delayed tests with an im-

TABLE 1

RECALL PROBABILITY AS A FUNCTION OF RETENTION INTERVAL FOR EXP. I AND II

Number of Prior Pairs	Exp.	Retention Interval (Sec.)				
		0	2	4	6	10
0	I	.984	.825	.484	.400	.350
	II	.986	.889	.430	.430	—
1	I	.975	.700	.378	.231	.259
	II	.967	.782	.352	.296	—
2	I	.959	.647	.325	.219	.209
	II	.907	.685	.426	.162	—
3	I	.944	.600	.300	.275	.262
4	II	.884	.694	.306	.273	—
8	II	.898	.625	.204	.227	—

TABLE 2

ANALYSES OF VARIANCE OF NUMBER OF CORRECT RECALLS IN EXP. I AND II

Source	Exp. I			Exp. II		
	df	*MS*	*F*	*df*	*MS*	*F*
Retention interval (t)	4	1893	253***	3	1264	255***
Prior pairs (PP)	3	103.7	34.3***	4	57.0	9.56***
*S*s	19	48.2		17	29.1	
t × *S*	76	7.49		51	4.96	
PP × *S*	57	3.02		68	5.96	
PP × t	12	8.25	2.84**	12	8.08	10.36***
PP × t × *S*	228	2.91		204	0.78	

** $p < .01$.
*** $p < .001$.

recent study by Conrad (1960), who found that serial order intrusions decreased as delay increased although the number of messages correctly recalled did not change. It was hoped that prior list intrusions might vary with time interval between lists, and these intrusions would be one measure of possible PI effects. Also, PI effects might vary as a function of time interval between prior and original learning.

All lists were six pairs long. With a recall interval of 15 sec., a "Ready" period of approximately 5 sec., and delay periods of 0, 10, or 25 sec., the total time that elapsed between the presentation of the last pair in one list and the first pair in the next list was either 20, 30, or 45 sec. With each delay period occurring equally often before each retention interval one replication required 18 lists (plus 1 extra list as a "starter" for each session).

The experiment was conducted in three sets of 4 replications each for a total of 12 replications (216 lists) in all. The order of presentation was randomized within each replication. The *S*s ($N = 23$) were tested in one afternoon session of 4 hr.; there was a 10-min. break after the first set of 4 replications and a 20-min. break after the second set of 4 replications.

The lists and strip films of Exp. III were used here, as was the group "memory drum" already described. No practice list was given, and the first list was: ORGAN-LOG; CONVENIENT-TITLE; STERN-CHECK; SCHOOLHOUSE-CANDIDATE; BELL-I'VE; HAST-CREATE; BELL? Thus, there was a 2-sec. retention interval and the correct response (i.e., I'VE) was given by 10 of 23 *S*s. In other respects the procedure was the same as that of Exp. III.

Experiment VI.—In Exp. I–V each *S* was tested repeatedly under all conditions. With repeated testing of the same *S*s PI could accumulate across lists and so contribute to the rapid forgetting. Experiment VI was conducted specifically to test this possibility, and naive *S*s were used throughout. That is, each *S* was tested once and only once, and there were different *S*s at each retention interval.

The A-B pairs used in this experiment had all been used as critical pairs in Exp. III, had appeared at either the 6-sec. or 8-sec. retention interval (where the curve is essentially flat; see Table 5), and were all at the median value (correctly recalled by 4 of 16 *S*s) for number of correct recalls for pairs at these two retention intervals. Thus, it seems safe to assume that the A-B pairs used in this experiment were of average difficulty.

In this experiment the lists were six pairs long and the presentation rate was 2 sec/pair. With a different group of naive *S*s at each retention interval, 6 groups of *S*s would be necessary for one replication. For greater generality there were two replications, or 12 different groups of naive *S*s in all. The specific lists used in each of the two replications are shown in Table 7. Each list was always presented in the order shown (i.e., from HEALTHY-BELT to SPREAD-HE), and the retention interval corresponds to the critical pair in the same row (i.e., when HEALTHY-BELT was the critical pair the retention interval was 10 sec.). In all there were 296 *S*s; group testing was used, and the size of each group is given in Table 7 as the denominator under proportion correct for each replication.

The procedure was carefully explained to each group before the list was presented. In effect, they were told that they would be shown a list of six pairs and tested for the recall of the second word of one of the pairs. The pairs were hand-written on heavy cardboard using a marking pen and letters 4 in. high. The cover card had written on it, CAN YOU SEE THIS? There seemed to be no trouble with legibility.

TABLE 7

LISTS AND RESULTS OF EXP. VI

Retention Interval (Sec.)	Replication A		Replication B		Weighted Average
	List	Proportion Correct	List	Proportion Correct	
10	HEALTHY-BELT	15/17	DELIGHTFUL-IMPORT	28/52	.62
8	SETTLE-LOUDLY	4/23	GAY-FACTOR	4/23	.17
6	HIGHWAY-LEAN	8/36	DESIRE-WHERE	15/37	.32
4	NOBODY-BURY	8/22	EARN-OPERATOR	2/16	.26
2	READING-ANGER	2/16	FEDERAL-DUKE	5/22	.18
0	SPREAD-HE	16/16	INTELLIGENT-HEAVY	13/16	.91

Results

The results of Exp. V were very clear-cut; there were no differences in retention as a function of intertrial interval and, for all practical purposes, there were no extralist intrusions from prior lists.

When summed over all six retention intervals the recall probabilities were .431, .444, and .429 for intertrial intervals of 15, 25, and 40 sec., respectively. An analysis of variance of number of correct recalls gave an F value less than 1.00. For intrusions, an analysis of the first 12 replications (72 lists) showed that, out of a total of 946 errors (any noncorrect recall), there were only 16 extralist intrusions from prior lists. Since these intrusions from prior lists accounted for less than 2% of the total errors, there seemed little justification for carrying the analysis any further.

If PI effects do vary as a function of time interval between prior and original learning there is no indication of any such effect occurring in Exp. V. Actually, if prior-list intrusions are in fact an index of PI one would have to conclude that there was little or no PI occurring in this situation. At least the results of Exp. V give no indication that any massive PI is accumulating to produce the rapid forgetting that does occur.

The main results of Exp. VI are shown in Table 7. The numerator of the fraction under proportion correct gives the number of *S*s who correctly recalled the critical pair at each retention interval. Thus, at the 10-sec. interval, 15 out of 17 *S*s correctly recalled BELT when shown HEALTHY as the cue for recall. Replications A and B are combined into a weighted average in the last column of Table 7.

The weighted average probably represents the best overall picture of the results of Exp. VI. First, it should be noted that the degree of original learning (0-sec. retention interval) seems to be just as high with naive *S*s as with *S*s who have been tested with hundreds of lists. That is, the weighted average for the 0-sec. retention interval was .91; the comparable value for Exp. III was .901 ± .085.

Second, the forgetting seems to be even more precipitous with naive *S*s than with practiced *S*s. That is, at the 2-sec. retention interval the weighted average was only .18; thus, in 2 sec. recall probability fell from .91 to .18. Judging by the results at the 4-sec. and 6-sec. retention intervals the value of .18 is probably somewhat low, but even if it were doubled or tripled it would still be less than the comparable value for Exp. III (i.e., .647 ± .120).

Third, there seems to be a very pronounced primacy effect. That is, the probability that the first pair would be correctly recalled was quite high (.62), even though the retention interval was fairly long (10 sec.). The comparable value for Exp. III was .277 ± .128.

DISCUSSION

The present results clearly indicate that, under the conditions of these experiments, there is very rapid forgetting of single paired associates. As in a previous report (Murdock, 1961b), probability of recall drops rapidly and approaches an asymptotic value of approximately .25 within a 10-sec. period of time. Comparable results under somewhat different conditions have been reported by Peterson, Saltzman, Hillner, and Land (1962).

In view of the importance generally assigned to PI as a cause of forgetting, its role in the present studies should be carefully assessed. First, however, we would like to suggest a distinction between two different types of PI that could operate in the short-term retention of single items. The first type is interlist PI, interference resulting from learning items or pairs in previous lists. The second type is intralist PI, interference resulting from learning items or pairs which precede the critical item or pair in the same list. Presumably interlist PI would gradually increase with stage of practice, whereas intralist PI would simply depend upon the number of prior items or prior pairs. Interlist and intralist PI would be independent factors although, of course, they could interact.

The stress on PI as a basic cause of forgetting in long-term memory stems from the finding that retention becomes poorer as stage of practice increases (Underwood, 1957, Fig. 3, p. 53). Recently the same effect has been suggested for the short-term retention of single items (Keppel & Underwood, 1962). For the present experiments this interpretation would demand that interlist PI increase with stage of practice.

The present experiments give no indication of a gradual build-up of interlist PI with stage of practice. To summarize, the evidence is as follows: First, over the four sessions performance generally improved, and the improvement was significant at the .01 level for Exp. I and II and for 1 *S* (TE) of Exp. IV. Second, there was no effect of intertrial interval in Exp. V. If interlist PI was operating it might reasonably be expected to vary over time. Third, for all practical purposes there were no prior-list intrusions in Exp. V, and prior-list intrusions are conventionally regarded as one index of PI. Fourth, the recall probabilities obtained on the practice lists and/or the first test list of Exp. I, II, III, and V were never above the overall results of the comparable condition within each experiment; in general, they were below. Fifth, the forgetting of naive *S*s in Exp. VI was, if anything, more rapid than that of practiced *S*s. Therefore, it would seem that interlist PI is not a significant factor in the present experiments.

The results of Exp. VI do suggest that interlist PI may have one noticeable effect. This experiment showed that, with naive *S*s, the recall probability for the first pair in the list was much higher than with practiced *S*s. If this finding holds up in further replications the obvious implication is that interlist PI reduces a primacy effect.[2]

While interlist PI apparently plays no role except to reduce a possible primacy effect, intralist PI does affect the recall of single paired associates. As the results of Exp. I and II indicated, recall probability did decrease significantly as number of prior pairs increased. While statistically significant, the magnitude of this effect was not very great. And, as has been pointed out, the confounding of number of prior pairs with retention interval did not seem to affect the results of Exp. III and IV to any great extent.

As Table 1 shows, degree of original learning (i.e., recall probability at 0-sec. retention interval) varies inversely with number of prior pairs. Analyses of variance of number of correct recalls showed that this effect was significant both in Exp. I, $F\ (3, 57) = 3.15, p < .05$, and in Exp. II, $F\ (4, 68) = 5.27, p = < .01$. In some cases there is probably no need to correct for these differences in degree of original learning. For instance, Murdock (1962a) has shown that, with fixed-length lists, the retention curve for the recall of A given B is essentially indistinguishable from the retention curve for the recall of B given A. Since in this experiment the main interest was in direction of recall there was no need to correct for possible differences in degree of original learning.

However, in other cases it might be desirable to make such a correction. To aid in this correction, the data from the first column of Table 1 were plotted, and it was found that they could be fairly well described by the Gompertz function

---

[2] If these results can be generalized to free recall, they may explain why Murdock (1962b) failed to replicate a primacy effect reported by Bousfield, Whitmarsh, and Esterson (1958). In the former study the same *S*s were repeatedly tested, while in the latter study *S*s were tested once at each length list.

$p = 1.000 - (.110)\,(.129)^{.578^{n}}$, where p is recall probability and n is number of prior pairs. In effect, this equation gives degree of original learning as a function of number of prior pairs. For example, with five prior pairs $p = .904$; as Table 5 shows, with $n = 5$ (all lists were six pairs long) the recall probability for the 0-sec. retention interval in Exp. III was .901.

The corrected retention loss for Exp. III was .266, .571, .666, .694, and .709 for retention intervals of 2, 4, 6, 8, and 10 sec., respectively. That is, corrected retention loss is retention loss (as shown, for instance, in Table 3) when degree of original learning is corrected for number of prior pairs by the formula given above. The corrected retention loss data clearly suggest that the curve of the forgetting of single paired associates rapidly approaches an asymptote of about .70 within a 10-sec. retention interval.

It would seem, then, that intralist PI does have some effect on retention and a correction for degree of original learning may be in order when a high degree of precision is required. However, even intralist PI did not seem to operate in the expected manner. That is, it is generally accepted that PI effects increase over time (Slamecka & Ceraso, 1960). However, as Table 3 shows, with single paired associates PI effects appear to decrease over time. That is, the effect of number of prior pairs was greatest at the shortest retention interval, and at the longer retention intervals the effect was so irregular as to be questionable.[3]

In conclusion, it would appear that the generally accepted principles of long-term retention are not too helpful in explaining the present results. Interlist PI effects were notable by their absence while intralist PI effects decreased rather than increased over time. Just what does account for the rapid forgetting is far from clear. It may be that, with the presentation of each pair, a perceptual trace of some kind is formed which rapidly decays over time. Or, perhaps, during the retention interval there is some sort of spontaneous recovery of pre-experimentally acquired associations which serve to interfere with the B response to the critical A stimulus. Obviously there are other possibilities, but further work is necessary to provide an adequate explanation of this phenomenon.

REFERENCES

BOUSFIELD, W. A., WHITMARSH, G. A., & ESTERSON, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. *J. gen. Psychol.*, 1958, **59**, 255–262.

CONRAD, R. Serial order intrusions in immediate memory. *Brit. J. Psychol.*, 1960, **51**, 45–48.

KEPPEL, G., & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single items. *J. verbal Learn. verbal Behav.*, 1962, **1**, 153–161.

MURDOCK, B. B., JR. The retention of individual items. *J. exp. Psychol.*, 1961, **62**, 618–625. (a)

MURDOCK, B. B., JR. Short-term retention of single paired associates. *Psychol. Rep.*, 1961, **8**, 280. (b)

MURDOCK, B. B., JR. Direction of recall in short-term memory. *J. verbal Learn. verbal Behav.*, 1962, **1**, 119–124. (a)

MURDOCK, B. B., JR. The serial position effect of free recall. *J. exp. Psychol.*, 1962, **64**, 482–488. (b)

PETERSON, L. R., SALTZMAN, D., HILLNER, K., & LAND, V. Recency and frequency in paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 396–403.

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

(Received June 1, 1962)

---

[3] In a previous study (Murdock, 1961a) intralist PI appeared to have a U shaped effect; no such curvilinearity was apparent here.

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 444–450

EFFECT OF CONTINUED NONREINFORCEMENT ON THE FRUSTRATION EFFECT[1]

JAMES H. McHOSE[2]

State University of Iowa

A total of 72 preshift and 144 postshift trials in the double alley runway was given to 48 female hooded rats randomly assigned to the following 3 groups: Groups 100-50, 100-0, and 0-0, designated according to preshift and postshift percentage reinforcement in the 1st goal box (G_1). Following the introduction of nonreinforcement, 2nd alley starting speeds for both Group 100-0 and Group 100-50 on G_1 nonreinforced trials were enhanced. The facilitative effects of nonreinforcement dissipated for Group 100-0, but persisted for Group 100-50. Group 100-50 1st alley starting speeds increased while Group 100-0 speeds decreased following the introduction of nonreinforcement. The results were interpreted as supportive of the frustration theory assumption that the parameters of frustration are dependent upon the strength of reward expectancy in the 1st alley.

In recent years a series of investigations using the double runway apparatus have reported a frustration effect (FE) resulting from nonreinforcement of a previously reinforced response (Amsel & Hancock, 1957; Amsel & Roussel, 1952; Wagner, 1957).

In the context of behavior theory of the Hull-Spence variety, the FE has been interpreted (Amsel, 1958) as reflecting an increase in the drive level of the organism in the second alley (A_2) following nonreinforcement in the first goal box (G_1). According to Amsel this increased drive level is the result of a hypothetical emotional (anger, frustration) response that occurs with nonreinforcement. The theory further assumes that the occurrence and intensity of this frustration response (R_F) is in turn dependent upon the prior development in the instrumental chain of an expectation of reward or, in terms of Spence's (1956) theory, of a conditioned fractional anticipatory goal response (r_g), in the first alley (A_1).

The present study is primarily concerned with the aspect of this theoretical interpretation which assumes that the parameters of R_F are based upon those of the conditioned r_g in the first runway. If after training in the double runway situation, reinforcement in G_1 should be discontinued, it would be expected that at first R_F should occur in G_1 and, consequently, speed of running in the second alley (A_2) should be augmented compared with a control condition in which there was no frustration in G_1. But with discontinuance of reinforcement in G_1, it would be expected that the strength of the fractional anticipatory goal response in A_1 would eventually become weakened and, to the extent that A_1 and A_2 are dissimilar, tend to approach zero strength. With this decrement in the strength of the conditioned r_g, it would be expected that FE would also tend to decrease.

---

[1] This experiment represents a portion of a dissertation submitted to the Psychology Department of the State University of Iowa in partial fulfillment of the requirements for the PhD degree. The author wishes to acknowledge his indebtedness to Kenneth W. Spence for advice and assistance throughout the course of the investigation. The experiment was conducted during the author's tenure as a National Science Foundation Cooperative Graduate Fellow.

[2] Now at Southern Illinois University.

In order to test this implication of the theory, three groups of *S*s were employed. Two of the groups were reinforced on 100% of the trials in the first and second goal boxes in the first phase of training. In the second, postshift phase of the experiment one of these groups (Group 100-0) received no reinforcement in G_1 while the other (Group 100-50) received reinforcement in G_1 on a random 50% of the trials. Reinforcement in the second goal box was continued on all trials for both groups during the postshift phase. A third group (Group 0-0) received no reinforcement in G_1 throughout the course of the investigation and was always reinforced in the second goal box. This group, as in Wagner's (1957) study, served as a control to determine whether nonreinforcement in G_1, after a history of previous reinforcements, led to augmentation of the subsequent performance in A_2.

According to the theoretical interpretation outlined above, the initial postshift performance in A_2 of the group shifted to 0% reinforcement in G_1 should exhibit an increase comparable to that shown by Group 100-50 on trials on which no reinforcement was given in G_1. With the expected reduction in strength of r_g in A_1 as a consequence of the discontinuation of reinforcement in G_1, the performance of Group 100-0 should subsequently decrease to a level of that of Group 100-50 on those trials on which it received reinforcement in G_1 and, presumably, also to a level approximating that of Group 0-0 throughout this same period of training.

Performance in the first alley was also of interest, since it would seem reasonable to assume that any decrease in the strength of the conditioned r_g in A_1 would result in a decrease in the performance level in that alley. Thus the 100-0 group's performance level in that alley would be expected to exhibit an eventual decrease in the postshift phase of the experiment.

METHOD

Subjects.—Forty-eight experimentally naive female hooded rats from the colony maintained by the Psychology Department of the State University of Iowa served as *S*s. Their ages ranged from 83 to 93 days at the beginning of the experiment.

Apparatus.—The double alley apparatus consisted of an L shaped double runway. The first goal box served as the second start box.

The height of the apparatus was 4 in. and the width 3 in. at all points with the exception of the first start box, which was 2 in. wide. The length of the various segments of the apparatus were as follows: the first start box (S_1), 12 in.; the first alley (A_1), 33 in.; the first goal box (G_1 or S_2), 15 in.; the second alley (A_2), 36 in.; the second goal box (G_2), 12 in. Sections S_1, A_1, and G_1 were painted flat black and had a black ribbed plastic floor, while Sections A_2 and G_2 were painted white and had a smooth white floor. Sections S_1 and G_2 had removable Plexiglas tops; the remaining segments of the apparatus were covered with plate glass. A small metal cup mounted on the wall of G_1 served as the foodcup in that goal box and a small wooden block with an indentation for food pellets served as the foodcup in G_2.

The apparatus had four doors: a dropping start door between S_1 and A_1, a guillotine start door between S_2 and A_2, a sliding door between A_1 and G_1, and a guillotine door between A_2 and G_2. Dropping the first start door activated a Hunter Klockkounter which stopped when *S* interrupted a photocell beam located 6 in. from S_1. Raising the second start door activated a second counter which stopped when *S* interrupted a second photocell beam located 6 in. from S_2. Photocell beams were passed through infrared filters.

Procedure.—Fifteen days prior to the first training day, *S*s were placed on a 23-hr. food deprivation schedule employed throughout the experimental period. Feeding was always conducted in the home cages and, with the exception noted below, always consisted of 1 hr. access to Purina lab chow. On the 2 days immediately preceding the first training day, all *S*s were given 30 min. feeding with Noyes pellets and 30 min. feeding with Purina lab chow. During both the habituation and experimental period, water was available at all times in both home and carrying cages.

The experimental procedure consisted of two stages: a 72-trial acquisition phase and a 144-trial postshift period. In the acquisition phase 32 *S*s received two .045-gm. Noyes pellets in both G_1 and G_2 and 16 *S*s received no pellets in G_1 and two pellets in G_2. Following Trial 72, the group that had been rewarded in G_1 was subdivided into two groups of 16 *S*s each, matched according to mean A_2 starting speeds over Trials 57–72. One of these groups (Group 100-0) received no reward in G_1 during the postshift phase; the other group (Group 100-50) received two pellets in G_1 on half the postshift trials and no reward on half the trials according to the following repeating cycle: FRFR, RFRF, RFFR, FRRF, RRFF, FFRR where R and F designate rewarded and nonrewarded (frustrated) trials, respectively. Those *S*s that received no pellets in G_1 during acquisition continued to receive no reward during the postshift phase of the experiment. Subjects received four trials per day with the exception of the first 2 days of training, on which *S*s received two trials. The experiment was conducted in two replications.

A trial began with the introduction of *S* into the first start box. The first start door was dropped after *S* maintained an orientation toward it for 3 sec. When *S* entered G_1, the sliding door separating A_1 and G_1 was closed. The *S* either received no reward or two pellets in G_1, according to the group to which *S* was assigned. After *S* had eaten, *E* counted 1 sec. and raised the second start door. (In the case of nonrewarded trials, *E* counted 4 sec. before raising the door.) When *S* had entered G_2, the guillotine door separating A_2 and G_2 was closed and *S* was removed immediately after eating. All *S*s were fed at the end of the last experimental trial for the last *S* of the day.

RESULTS AND DISCUSSION

Figure 1 presents group mean Alley 2 starting speeds as a function of trial blocks. The vertical line after Trial Block 9 indicates the point at

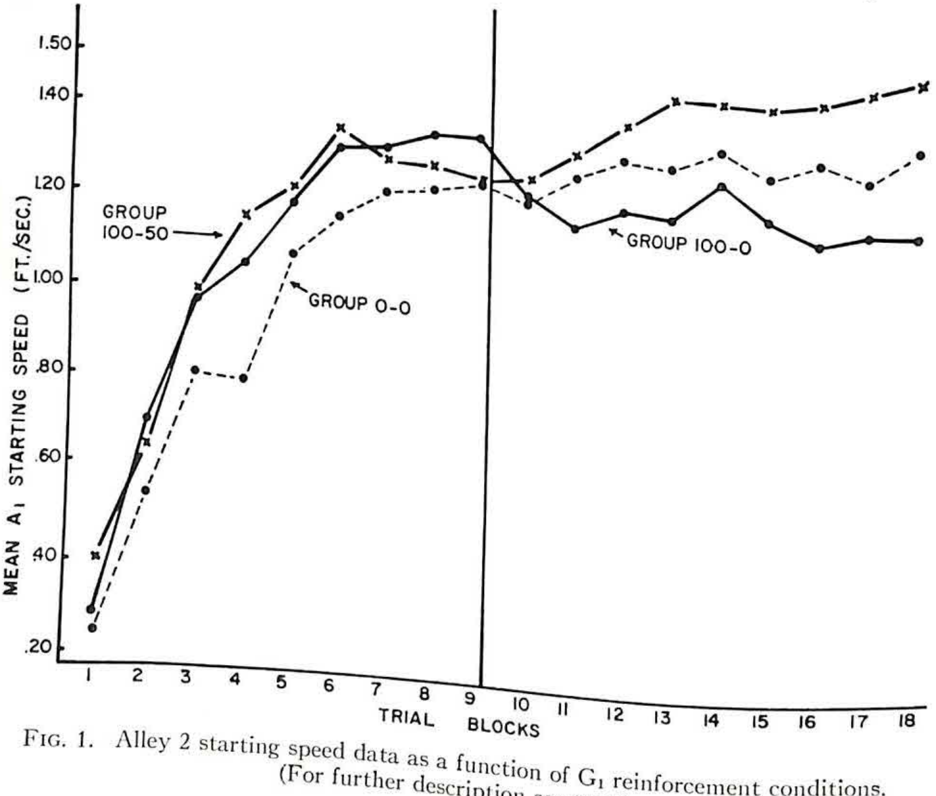


FIG. 1. Alley 2 starting speed data as a function of G_1 reinforcement conditions. (For further description see text.)

which the reinforcement schedule in G_1 was changed for Groups 100-50 and 100-0. Each preshift block consists of 8 trials. The data for the postshift trials were separated into blocks of 16 trials. The means of each of these blocks were then plotted for Groups 100-0 and 0-0. For Group 100-50, however, mean speeds for the 8 trials on which *S*s were rewarded in G_1 (R) and the 8 trials on which they received no reward (F) were plotted separately.

Inspection of Fig. 1 indicates that the data for Group 100-50 not only exhibited the FE following the introduction of nonreinforcement in G_1, but that this effect persisted throughout the 144 postshift trials. A Treatment × Treatment × *S* analysis of variance (all variance analyses used are identified according to Lindquist's, 1953, terminology) over the first (10–11) and last (17–18) two trial blocks of this phase indicated that the FE was significant at the .001 level, $F\ (1, 15) = 36.4$. A Treatment × *S*s analysis of this group's performance on the last two preshift blocks (8–9) and its subsequent performance on the first two postshift trial blocks of frustrated trials revealed an *F* ratio significant beyond the .025 level, $F\ (1, 15) = 7.96$.

Demonstration of FE is in accord with results of previous investigators (Amsel & Hancock, 1957; Amsel & Roussel, 1952; Wagner, 1957). The higher performance level following nonreinforcement in G_1 supports the position that nonreinforcement of a previously reinforced response results in an increase in drive level (*D*). The increase in *D* is interpreted as due to frustration in G_1. Previous investigations were supplemented in that the persistence of FE over 144 trials was demonstrated. Previous investigators had not extended the postshift period beyond 54 trials (Wagner, 1957). If, as stated in frustration theory, FE depends on the strength of reward expectancy in G_1, then this persistence would be expected since Group 100-50 was fed in G_1 on a random half of the postshift trials, thereby maintaining the conditioned r_g.

Turning next to the group in which none of the trials in the postshift period were rewarded in G_1 (Group 100-0), it may be noted that the performance of this group, like that of Group 100-50 on frustrated trials, showed a marked increase on the first postshift block of 16 trials. Unlike the 100-50 F performance curve, however, that for Group 100-0 gradually dropped to the level of performance exhibited by Group 100-50 on G_1 rewarded trials. Analysis of variance of the performance data represented in Curves 100-50 F and 100-0 over Trial Blocks 10–11 and 17–18 revealed a significant ($p < .01$) Trials × Groups interaction, $F\ (1, 30) = 12.09$. A simple variance analysis of the last two trial blocks (17–18) indicated the difference between groups was significant ($p < .05$) at these points, $F\ (1, 31) = 5.45$. Analogous analysis of Group 100-0's performance with that for Group 100-50 on R trials provided a Trials × Groups interaction *F* ratio significant at the .05 level, $F\ (1, 30) = 6.34$. However the difference between these performances at Trial Blocks 10–11 fell short of significance ($p > .10$). The obtained interaction nevertheless suggests that there was a difference between groups either at the first (10–11) or second (17–18) trial blocks.

A Treatment × *S* analysis of Group 100-0 late preshift (8–9) and early postshift (10–11) performance also yielded an *F* ratio short of significance ($p > .10$).

Neither the comparison of 100-0 postshift performance to that of

100-50 on R trials nor the analysis of preshift and postshift performance of Group 100-0 offers reliable evidence that nonreward facilitated 100-0 performance. However, the comparisons of 100-0 and 100-50 data suggest such an effect, as does the significant Trials × Groups interaction obtained in the analysis of 100-0 and 100-50 R data.

If the evidence of enhanced 100-0 performance is accepted, the analyses clearly suggest that the facilitative effects of nonreinforcement dissipate over the postshift period. These results suggest that the frustrative effects of nonreward are indeed dependent on the expectancy of reward in the first goal box ($S_1 - r_g$). The postshift phase consisted of extinction of r_g for Group 100-0. The extinction of r_g, as implied in frustration theory, resulted in dissipation of the frustration response (R_F) and, consequently, a decrease in drive level.

As may be seen in Fig. 1, the control (0-0) group exhibited postshift performance above that for Group 100-50 on R trials, but below that for Group 100-50 on F trials. Analyses of variance over early (10–11) and late (17–18) postshift trials indicated no significant differences between the control group performance and either the 100-50 F or 100-50 R performance levels. These analyses also revealed no significant Trials × Groups interactions over these trial blocks.

The present control group's performance is difficult to interpret. It differs from Wagner's (1957) finding in that in the latter study the postshift performance of the group that never received reward in G_1 was at the same level as the performance on nonreinforced trials of the group rewarded on 50% of the trials in G_1, i.e., 100-50 R. One possibility is that the asymptotic performance of the present group was, because of a sampling difference, somewhat above that of the two experimental groups. The fact that the performance of the control (0-0) group did not differ significantly at the end of training (Blocks 17–18) from the performance of the experimental groups does not support the possible interpretation that the latter groups' performance was depressed on rewarded trials. Also the fact that the final postshift performance levels of the 100-0 group and the 100-50 group on R trials did not differ appreciably from their final preshift performance levels precludes this interpretation. Regardless of the interpretation of the present control group performance, it should be noted that the facilitative effects of nonreward were not demonstrated insofar as the comparisons are based on the 0-0 group, but only within the experimental *S*s.

Figure 2 presents group mean Alley 1 starting speed data for the three groups as a function of trial blocks. Each preshift trial block consists of 8 trials and each postshift block of 16 trials. The vertical line represents the point at which nonreinforcement was introduced for Groups 100-0 and 100-50. As may be seen, all groups appear to have stabilized at approximately the same asymptote by the end of the preshift period. Simple analysis of variance indicated that there were no significant differences between groups over Trial Blocks 8–9.

Considering next the postshift data, Fig. 2 shows that the 100-50 curve rises with continued training whereas the 100-0 curve decreases. Analyses of variance were performed over Trial Blocks 8–9 and 18 to compare Group 0-0 with Group 100-50 and with Group 100-0. These analyses yielded Trials × Groups interaction *F* ratios significant at the .05 and .001 levels, respectively, $F\ (1,\ 28) = 4.66$ and

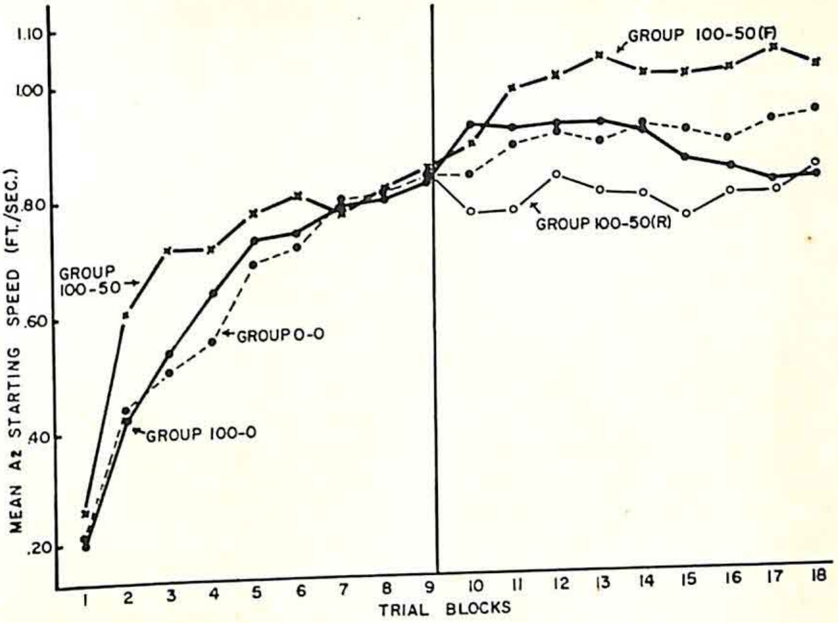


FIG. 2. Alley 1 starting speed data as a function of G_1 reinforcement conditions. (For further description see text.)

$F(1, 28) = 13.89$, indicating that the curves for Groups 100-50 and 100-0 diverged from that for Group 0-0 during the postshift period. Treatment $\times S$ analyses over these trial blocks revealed significant trial effects for the 100-50 and 100-0 groups, $F(1, 28) = 23.34$ and $F(1, 28) = 12.16$, but not for the 0-0 group, although the trial effect for the latter group approached significance ($p > .05$).

An interpretation of the Alley 1 postshift data in terms of frustration theory is fairly simple. The introduction of continual nonreward for Group 100-0 would lead to the extinction of the conditioned r_g in the first alley and, since r_g is assumed to facilitate performance, this would lead to a performance decrease.

The shift in G_1 percentage reinforcement for the 100-50 group, however, would lead to conditioned frustration in Alley 1, and the consequent increase in drive level would result in enhanced performance.

Assuming that Group 100-50 was at asymptote prior to the shift in G_1 percentage reinforcement, the significant performance increase may be taken as evidence supporting the conditioned frustration mechanism. It should be noted that Wagner (1957) has reported this same phenomenon. Moreover, this phenomenon is similar to the finding that 50% reinforcement results in higher asymptotic performance in the single alley situation than does 100% reinforcement (cf., Goodrich, 1959; Wagner, 1961; Weinstock, 1954). This latter phenomenon has also been interpreted (Spence, 1960) in terms of the development of conditioned frustration on the 50% reinforcement schedule.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A., & HANCOCK, W. Motivational properties of frustration: III. Relation of frustration to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–64.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

WAGNER, A. R. Motivational effects of nonreinforcement as a function of the reinforcement schedule. Unpublished master's thesis, State University of Iowa, 1957.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. Unpublished doctoral dissertation, Indiana University, 1954.

(Received June 21, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 451–454

PREFERENCE FOR BAR PRESSING OVER "FREELOADING" AS A FUNCTION OF NUMBER OF REWARDED PRESSES[1]

GLEN D. JENSEN

DePaul University

200 rats were given a choice between eating pellets from a dish attached to the floor of a Skinner box or pressing a bar to earn pellets after 40, 80, 160, 320, 640, or 1280 rewarded presses. All but 1 of the rats left the food dish and resumed pressing the bar at some point during a 40-min. choice period. The mean percentage of all pellets eaten during the choice period that were earned by bar pressing was shown to be an increasing function of the number of rewarded presses made prior to the choice period.

In their efforts to avoid anthropomorphizing, psychologists seem to have ignored the possibility that common operants such as maze running or bar pressing hold varying amounts of *intrinsic appeal* for the rat. *Intrinsic appeal* may be thought of as a pleasant emotional state experienced while performing the operant, but an operational definition is needed. We will say that an operant has intrinsic appeal for a species if members of that species show a preference for obtaining food by using that operant as opposed to using a less-effortful operant or a better-established operant that would lead to an equal or greater amount of food per unit of time. It is felt that these operations are sufficient to distinguish intrinsic appeal from *operant level* which is commonly defined, in the Skinner-box situation, as rate of pressing prior to the first reward. It may well be, however, that future research will prove the concepts to be directly related. It seems likely that some situations which show high intrinsic appeal will also be characterized by high operant levels, but this is strictly an empirical question.

This paper reports an attempt to demonstrate that bar pressing has an appreciable amount of intrinsic appeal for the rat. Further, an attempt is made to demonstrate how intrinsic appeal varies over a rather wide range of rewarded presses made prior to the test of intrinsic appeal.

METHOD

Subjects.—The *S*s that completed the experiment were 200 male albino rats from 68 to 140 days of age. Two additional *S*s were discarded because of failure to learn the bar pressing response and 3 because of illness. The first 140 *S*s were randomly assigned to seven experimental conditions subject to the restriction that each group contain 20 *S*s. The last 60 *S*s were randomly assigned to three experimental conditions that were strategically selected replications of three of the previously administered conditions.

Apparatus.—Three Skinner boxes comparable to the one described by Jensen (1961) were used.

Procedure.—For a minimum of 10 days prior to the Dispenser Training Day each *S* was fed 10 gm. of finely ground Purina lab checkers 1 hr. after the time of day when *S* was later to be run.

On the Dispenser Training Day *S* was placed in the Skinner box for 25 min. During the 25 min., the pellet dispenser automatically delivered a 45-mg. pellet every 30 sec. until a total of 50 pellets had been delivered. The bar was covered by a tin shield during this session. At the end of the session, *E* returned the rat to his home cage and fed him 10 gm. of ground checkers minus the weight of the 50 pellets $\frac{1}{2}$ hr. later.

Twenty-four hours after the Dispenser Training Day, each *S* was placed in the

---

[1] The research was supported by Grants M-2834 and M-4494 from the United States Public Health Service.

Skinner box and allowed to make 40 rewarded presses. All *S*s but those in Group 40 were then returned to their home cages and fed 10 gm. minus the weight of 40 pellets at the end of the hour. For an *S* in Group 40, the Skinner box was opened after the fortieth press and the power was turned off. The *E* then inserted a tin cup containing 250 45-mg. pellets. The cup was 2½ in. in diameter and 1¼ in. deep. It was identical to the cup from which *S*s had eaten in their home cages for anywhere from 10 to 90 days. The cup was securely fastened to the floor of the test chamber in the corner furthermost from the bar. After *S* had eaten at least two pellets from the cup and while he was still facing the cup, *E* gently reclosed the Skinner box and turned the power back on, so that *S* had a choice between eating pellets from the cup or by pressing the bar. The choice period lasted 40 min. Then *E* computed the percentage of all pellets eaten during the choice period that had been "earned" by bar pressing. To do so, it was necessary to dig out and count pellets that were left in the delivery tube of the pellet dispenser, in the food cup, and below the floor near the food cup or the pellet hopper. A funnel arrangement below the grid floor minimized danger of confusing pellets spilled near the bar from those dropped near the food cup.

The *S*s in Group 80 made 40 more rewarded presses on Day 2 and then were given a choice period identical to that of Group 40. Group 160A made 80 rewarded presses on Day 2 and then received a "40 plus choice" session on Day 3 (the same treatment that Group 40 received on Day 1 and Group 80 received on Day 2). Group 160B made 40 rewarded presses on each of Days 2 and 3 and then received "40 plus choice" treatment on Day 4. The *S*s in Group 320 made 120 rewarded presses on Days 2 and 3 and then received "40 plus choice" treatment on Day 4. The *S*s in Group 1,280 made 120 rewarded presses on Days 2 and 3, 160 rewarded presses during each of Days 4–9, and received "40 plus choice" treatment on Day 10. On all days except the choice day, *S*'s ration was adjusted for the number of pellets eaten while bar pressing. The conditions for Groups 40, 320, and 640 were replicated at a later date with 60 new *S*s.

RESULTS AND DISCUSSION

The proposition that the typical rat prefers bar pressing for food to "freeloading" from the food cup at some point prior to satiation is well indicated by the fact that only 1 *S* of the 200 run ate 100% of his pellets from the food cup. Also, 44% of the *S*s earned more than half of all pellets eaten during the choice period by bar pressing.

Figure 1 shows the relationship between preference for bar pressing over "freeloading" and number of rewarded bar presses prior to the choice period. The solid circles represent the original seven groups. Analysis of the scores from which these means were determined yielded a difference significant at the .01 level, $F\ (6, 133) = 4.49$. However, a subsequent Tukey[2] analysis for comparing individual means revealed that Group 1,280 differed significantly from all other groups at the .05 level or better, but no other comparisons were significant.

The open circles in Fig. 1 show the results of a partial replication. Group 40R differed from each of the other replicated groups at the .05 level, but the other groups were not significantly different from each other.

Group 160B which had an extra rewarded session relative to Group 160A showed (nonsignificantly) less preference for bar pressing than Group 160A. Consequently there is no indication that number of rewarded sessions as opposed to number of rewarded presses is a major determiner of preference for bar pressing.

Until better evidence is available, the equation shown in Fig. 1 will be assumed to summarize the relationship between the dependent and the independent variables of these experiments. The equation represents the best fit that can be obtained with a fairly simple exponential function

---

[2] J. W. Tukey, privately circulated monograph entitled "The Problem of Multiple Comparisons," 1953.

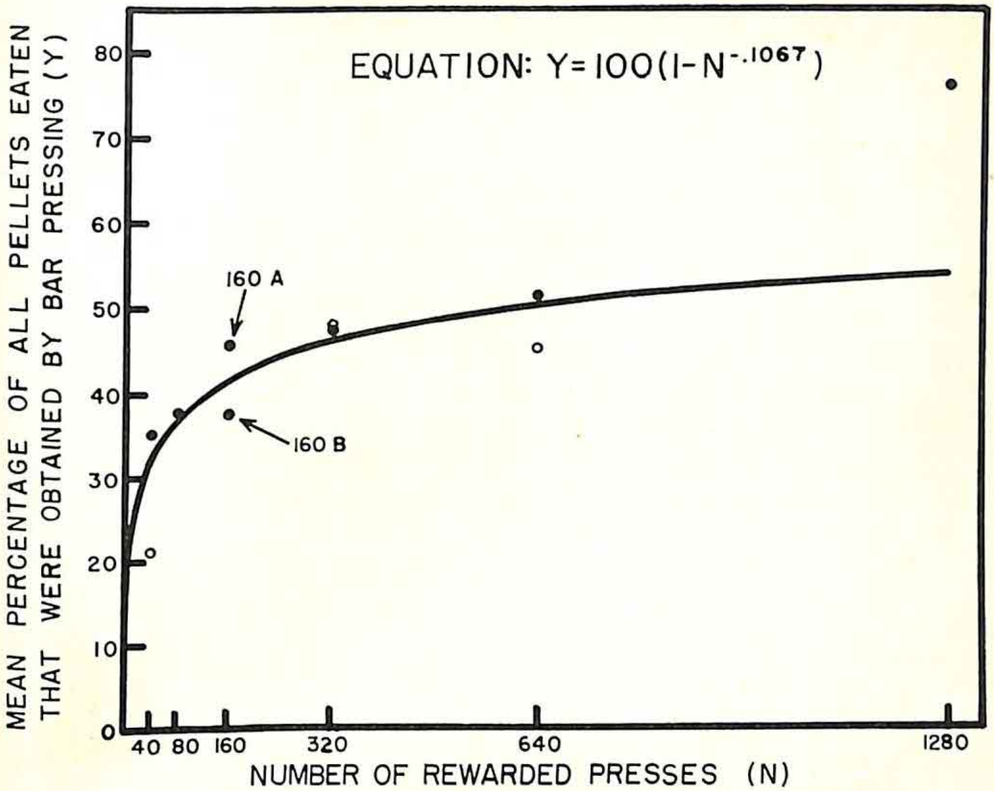


FIG. 1. Mean percentage of all pellets eaten during the 40-min. choice period that were earned by bar pressing as a function of number of previous rewards.

when all 10 points are given equal weight. The multiplier was rationally set at 100 because it is analytically true that rats can not obtain more than 100% of their pellets by bar pressing. The point for Group 1,280 is badly missed. This can be remedied only by using a considerably more complex equation. Future research may demand transition to such an equation; but for the present we will assume that the point for Group 1,280 is spuriously high.

An increasing function such as that shown in Fig. 1 could be predicted from Hullian theory in terms of competing habits. The habit of seeing food in a cup and eating it might be said to compete with the habit of seeing the bar and pressing it. It is true that the relative strength of the latter habit to the former habit increased between 40 and 1,280 rewarded presses. However, the real question for Hullian theorists would seem to be that of why the rats pressed the bar at all. Even for *S*s in Group 1,280 the response of approaching food to the characteristic stimuli of food and hunger had been reinforced many more times and with less delay than the response of pressing the bar to the characteristic stimuli of the bar. During each of the 1,280 rewarded presses the response of approaching food was reinforced, and with less delay than the response of pressing the bar. Also, the response of approaching food was rewarded many times outside of the Skinner box.

Prior to the present experiment about 100 rats had been given a similar choice situation except that the food cup was merely inserted without making sure that *S* had started eating and was facing the food cup as the box was reclosed. Under these conditions many *S*s continued pressing without ever eating a single pellet from the food dish. These operations left the door open for a Guthrian interpretation. That is, it could be said that the rat merely continued doing what he had last done, namely, pressing the bar. It would seem that the present operations make such an interpretation less plausible, since the last thing the *S* was doing before the onset of the choice period was eating from the dish.

It is unfortunate that no means were available for measuring the number of times that a rat switched from cup eating to bar pressing and the relative time spent at each activity. However, the *E* frequently observed through a peephole and was able to form a strong impression to the effect that once a rat started pressing the bar he did not return to the food cup. This may have theoretical implications, but the author declines comment for the present.

The need for a concept such as intrinsic appeal is evident in the data of Cotton, Lewis, and Metzger (1958). They found that rats run much faster during early trials in the 4-ft. alley of a T maze than in a 4-ft. runway. The constructor of mathematical models who tries to predict absolute running speeds in both runway and T maze will be forced to arbitrarily assign different rates of approach to the asymptote to the two situations unless some independent evidence is found to warrant different rate constants. The technique for measuring intrinsic appeal may provide such evidence. It is predicted that rats who are given a choice between eating pellets in the start box or running a T maze after (e.g.) 20 rewarded runs will make their first run after eating fewer pellets than rats that have been similarly treated in a runway. It was recognized that this prediction has little chance of success if the running operant does not have some minimal amount of intrinsic appeal. Four rats have been given a choice between eating from a full food cup in the start box or traversing a 4-ft. runway for a single pellet. Each *S* had received 20 rewarded runs prior to the choice. All 4 *S*s traversed the runway after eating less than half their daily ration from the food cup.

REFERENCES

COTTON, J. W., LEWIS, D. J., & METZGER, R. Running behavior as a function of apparatus and restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

JENSEN, G. D. Partial reinforcement effects (PREs) and inverse PREs determined by position of a nonrewarded block of responses. *J. exp. Psychol.*, 1961, **62**, 461–467.

(Received June 22, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 455–459

SECONDARY REINFORCEMENT IN CHILDREN AS A FUNCTION OF CONDITIONING ASSOCIATIONS, EXTINCTION PERCENTAGES, AND STIMULUS TYPES[1]

JEROME L. MYERS AND NANCY A. MYERS

University of Massachusetts

96 preschool children were conditioned and extinguished in a free operant situation. Extinction responses resulted in the presentation of either a light or a buzzer on either 20% or 100% of the trials. The extinction stimulus was 1 of 3 types: it was paired either with reward only during conditioning (PAS), or with reward and nonreward (MAS), or with nonreward (NAS) only. The PAS resulted in most extinction responses followed by the MAS and then the NAS. More responses resulted from 100% presentation of the PAS than under the other conditions. The light was more effective than the buzzer.

Myers and Myers (1962) recently obtained strong secondary reinforcement effects during extinction of a free operant response in children with continuous presentation of a buzzer which had previously appeared following 20% of the responses during conditioning. The *S*s who had been conditioned in the same manner but received the buzzer following only 20% of their extinction responses made significantly fewer responses than those *S*s receiving the buzzer following 100% of their extinction responses; however, they made significantly more responses than a group receiving 0% buzzer in extinction. Clear secondary reinforcement effects have not always been obtained with either rats or children. Furthermore, there is some support for a discrimination hypothesis of secondary reinforcement effects (Bitterman, Fedderson, & Tyler, 1953; Melching, 1954) which would predict that the group extinguished under 20% buzzer should have the highest extinction rate, followed by the no-buzzer group, followed by the 100% buzzer group. It is therefore important to isolate the factors operating in the Myers and Myers study. One such factor may be the way in which the buzzer was presented during conditioning, i.e., on all reinforced trials, but only on reinforced trials. This is suggested by two experiments. Saltzman (1949) found that an excellent discrimination between goal boxes in a U maze (no food present) could be established, if the correct box had previously been the goal box on rewarded runway trials, and had been absent on nonrewarded runs. More recently, Miller (1961) has reported that "informative" stimuli (stimuli paired only with reward) have stronger secondary reinforcement effects than "redundant" stimuli (stimuli present on both rewarded and nonrewarded conditioning trials).

The major purpose of the present study is to investigate the relative effectiveness of various percentages of *positive association stimuli* (stimuli presented only on rewarded conditioning trials), *mixed association stimuli* (stimuli presented on both rewarded and nonrewarded conditioning trials),

---

[1] This research was supported by funds from National Institute of Mental Health Grant M-2620(C-2).

and *negative association stimuli* (stimuli presented only on nonrewarded trials) presented during extinction of an operant response in children. In addition, data will be provided on the relative effectiveness of two physically different stimuli which have been paired with primary reinforcement.

METHOD

Apparatus.—The apparatus employed was a portable box designed to attract the interest of preschool children. On the front was painted a clown face, having red jewel-light eyes, a push-button nose, and a slot-tray mouth. M & M coated chocolate candy was dispensed through a tube to the mouth of the clown, while a ½-sec. buzz was heard from the interior of the box. The *E* had access to and operated three silent knife switches which allowed administration of the predetermined reinforcement. The number of responses was recorded on an electric magnetic counter mounted on the back of the box, out of *S*'s sight. The *E* recorded the number of responses made during each successive 30 sec. of extinction.

Subjects.—The *S*s were 96 boys and girls between the ages of 4-9 and 6-1 attending kindergartens in Northampton, Massachusetts.[2]

Procedure.—Each *S* was asked in the classroom if he would like to play a clown game, and if he acquiesced he was led to a small testing room. The clown was placed on a small table and *S* was instructed to sit at the small chair in front of it. No written instructions were read to *S*, but *E* standardized the verbal instructions as much as possible. Attention was called to the clown face, especially the nose. The *S*s were told that "something happens when you press his nose. Let's see what happens." The *E* pressed the clown's nose and thereby turned on the eye-lights, and the buzz, and received an M & M. The *S* was encouraged to try it also and was given one rewarded preliminary trial. He was then told he could stay and play the game "as long as you want." The *E* then sat down behind the table, facing the open back of the box and *S*.

Each *S* was reinforced according to a predetermined reinforcement schedule which delivered a total of 15 M & M candies. Immediately following the fifteenth reinforcement, the extinction period commenced; no candy reinforcement was administered, and each *S* was run until he stopped and indicated a desire to return to the classroom or until 5 min. had elapsed, at which time *E* terminated the session. One last candy was offered at the end of the extinction period.

Design.—Four boys and four girls were assigned randomly to each of 12 groups. One stimulus (the mixed association stimulus) was presented on each trial during conditioning for 8 groups. The other stimulus (the positive association stimulus) was presented only on the 20% of the trials when a candy was delivered. The 8 groups differed only with respect to extinction treatment, one pair of groups (100% PAS) received the positive association stimulus for every button press in extinction; one pair of groups (20% PAS) received the positive association stimulus on approximately every fifth response, as in conditioning; the third pair of groups (100% MAS) received the mixed association stimulus for every button press in extinction; and the fourth pair of groups (20% MAS) received the mixed association stimulus on approximately every fifth response. For one group of each pair the eye-lights were the positive association stimulus and the buzz the mixed association stimulus; for the other group the reverse was true. Four additional groups received 20% positive association stimulus plus candy reinforcement during conditioning trials, as did the first 8 groups, but differed from these groups in that they received the other stimulus also on only 20% of the conditioning trials, and furthermore this stimulus (the negative association stimulus) *never* occurred on those trials when candy was delivered. One pair of these groups (100% NAS) received the negative association stimulus for every button press in extinction; the other pair (20% NAS) received the negative association stimulus for every fifth extinction response. Again for 1 group of each pair the eye-lights were the positive association stimulus, the buzz the negative association stimulus; for the other group this was reversed.

RESULTS

Figure 1 presents the mean number of responses for each ½ min. of extinction for each of the 12 experimental groups. The positive association

---

[2] The authors wish to thank W. Barry, Superintendent of Schools, and Esther Wien, Elementary School Supervisor, Northampton, Massachusetts, for their helpful cooperation in providing facilities and *S*s for this study.

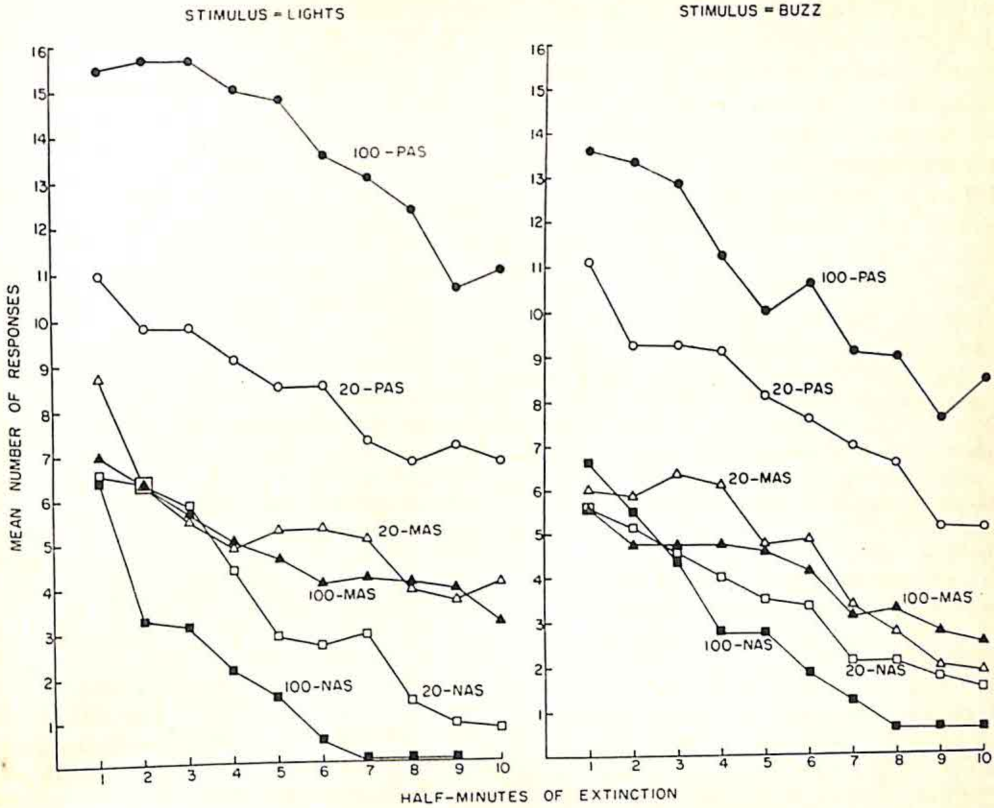


FIG. 1. Mean number of responses for successive half minutes of extinction.

stimulus (PAS) in extinction evokes the highest extinction response rate, the mixed association stimulus (MAS) yields fewer responses, and the negative association stimulus (NAS) yields the least responses under both percentages and both types of stimuli. The significant F ($p < .001$) for Schedules in Table 1 reflects these differences. Presentation of the stimulus following 100% of extinction responding yields more responses than 20% when the stimulus is a PAS, but less than 20% when the stimulus is either a MAS or NAS. This is supported by a significant F ($p < .001$) for Sc × P. Despite the interaction, the mean response rate of groups tested under 100% presentation is higher than that of groups tested under 20%. This is borne out by a significant F for Percentage ($p < .05$). The significant F ($p < .05$) for Stimulus reflects the higher response rate when the light is the extinction stimulus than when the buzzer is. The significant F ($p < .01$) for Sc × St is due to the fact that the separation of response rates under light and buzzer is greatest when the stimulus is a PAS, less when it is a MAS, and reverses slightly (i.e., the buzzer yields more responses) when the stimulus is a NAS.

The Trials main effect ($p < .001$) reflects the usual extinction occurrence; response rate drops over time. Response rate decreases slightly more rapidly when the extinction stimulus is a NAS than when it is a MAS, resulting in a significant Sc × T ($p < .05$) interaction.

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*
Total	959		
Between	95		
Schedules (Sc)	2	4,715.582	212.079***
Percentage (P)	1	152.801	6.872*
Stimulus (St)	1	134.254	6.038*
Sc × P	2	639.908	28.779***
Sc × St	2	112.619	5.065**
P × St	1	12.376	—
Sc × P × St	2	69.620	3.131*
Ss/Sc × P × St	84	22.235	—
Within	864		
Trials (T)	9	269.242	98.156***
Sc × T	18	5.592	2.039*
P × T	9	.035	—
St × T	9	1.811	—
Sc × P × T	18	4.535	1.653
Sc × St × T	18	3.484	1.270
P × St × T	9	2.156	—
Sc × P × St × T	18	3.260	1.188
Ss × T/Sc × P × St	756	2.743	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

A Duncan range test was used to compare the means of the six combinations of schedule and percentage. All means differed significantly from each other at the .01 level except 20-MAS and 100-MAS ($p > .05$).

DISCUSSION

In the present study the extinction stimulus was scheduled during conditioning in one of three ways: (*a*) it was presented on rewarded, and only on rewarded, trials (PAS); (*b*) it was presented on all trials, rewarded and nonrewarded (MAS); and (*c*) it was presented only on some of the nonrewarded trials (NAS). The PAS procedure yielded the most responses, the MAS next, and the NAS least. These results are consistent with the hypothesis that a stimulus must be discriminative (Dinsmoor, 1950; Schoenfeld, Antonitis, & Bersh, 1950) in order to be established as a secondary reinforcer. The PAS is a discriminative stimulus because its presence following the button press consistently signals the subsequent presence of candy, and its absence signals the absence of candy. The MAS, according to the discriminative stimulus hypothesis, should not become an effective secondary reinforcer since its presence does not permit *S* to discriminate between the subsequent presence and absence of reward. The discriminative stimulus hypothesis also accounts for the fact that the presentation of the NAS in extinction resulted in lower response rates than did the PAS or MAS. If a stimulus which consistently signals the presence of reward acquires secondary reinforcement properties, it follows logically that a stimulus which consistently signals the absence of reward will acquire negative secondary reinforcement properties; it becomes a punishment for the response. The discriminative stimulus hypothesis is also consistent with Miller's (1961) statement that a stimulus which is the only reliable predictor of the onset of food during conditioning is a strong secondary reinforcer.

The PAS evoked more extinction responses when presented on 100% of the extinction trials than when presented on 20% of those trials. Since the stimulus had appeared on 20% of the conditioning trials, the results contradict a "discrimination" (Bitterman et al., 1953) or "generalization" (Melching, 1954) hypothesis of secondary reinforcement, which would predict that the groups experiencing least change from conditioning to extinction would maintain the highest extinction response rates. However, the results of the present study are supported by previous findings that stronger secondary reinforcement is obtained under continuous than under partial presentation in extinction (Myers, 1960; Myers & Myers, 1962). There is presently no clear explanation of these differences in results, unless it be the dubious one that secondary reinforcers are formed with children and discriminations (between conditioning and extinction) are formed with rats.

The NAS evokes more responses when presented on 20% of the extinction trials than when presented on 100%. This is consistent with the results for the PAS, if one assumes that the NAS, through repeated presentations on nonrewarded conditioning trials, has become an aversive stimulus, or negative secondary reinforcer. In that case, fewer responses should occur, the more often it is presented. Within the framework of the above discussion, the fact that 20% MAS and 100% MAS do not differ significantly in effects on extinction rate suggests that a stimulus paired with these percentages of reward and nonreward is essentially neutral in its effects. This is supported by the fact that *S*s from the same population, tested with the same apparatus, a buzzer stimulus, and the same conditioning reinforcement schedule, with no stimulus in extinction, showed about the same response rate as the two MAS-buzzer groups did (Myers & Myers, 1962). Since 80% of the conditioning trials were nonrewarded, it follows that associations are formed more readily with reward than with nonreward.

The relative effectiveness of the two stimuli (buzzer and light) and the interaction of the stimulus with the conditioning schedule has important implications. The term "neutral stimulus," so often used in the literature, is inaccurate and misleading. Some of the apparent contradictions in the secondary reinforcement literature may simply be a function of the differences between clicks, buzzers, tones, lights, and black (or white) goal boxes.

REFERENCES

BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

DINSMOOR, J. A. A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *J. exp. Psychol.*, 1950, **40**, 458–472.

MELCHING, W. H. The acquired reward value of an intermittently presented neutral stimulus. *J. comp. physiol. Psychol.*, 1954, **47**, 370–374.

MILLER, N. E. Analytical studies of drive and reward. *Amer. Psychologist*, 1961, **16**, 739–754.

MYERS, N. A. Extinction following partial and continuous primary and secondary reinforcement. *J. exp. Psychol.*, 1960, **60**, 172–179.

MYERS, N. A., & MYERS, J. L. Effects of secondary reinforcement schedules in extinction on children's responding. *J. exp. Psychol.*, 1962, **64**, 586–588.

SALTZMAN, I. J. Maze learning in the absence of primary reinforcement: A study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, **42**, 161–173.

SCHOENFELD, W. N., ANTONITIS, J. J., & BERSH, P. J. A preliminary study of training conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 40–45.

(Received June 23, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 460–467

SHORT-TERM RETENTION AS A FUNCTION OF CONTEXTUAL CONSTRAINT[1]

KENNETH E. LLOYD AND WILLIAM A. JOHNSTON

Washington State University

Contextual constraint has improved recall in the classic memory span task. Here this variable was manipulated in a sequential memory task in which S continually received information, and, at unpredictable moments, was requested to recall some of it. In Experiment I 24 undergraduates continuously stored an average of 3 or 7 items and recalled them 2 at a time. This required many recall points which interrupted the continuity of contextual passages. The effect of varying contextual constraint was significant ($p < .01$), but small. In Experiment II interruptions were gradually reduced by requesting more and more items per recall point. Recall improved ($p < .01$) as constraint increased and as interruptions decreased; these variables interacted. Interruptions severely curtailed the benefits of context.

Several *E*s have recently studied short-term memory in a continuous, or sequential, task (Lloyd, Reid, & Feallock, 1960; Pollack, Johnson, & Knaff, 1959; Shepard & Teghtsoonian, 1961; Yntema & Mueser, 1960). In sequential memory tasks *E* presents *S* with a long series of items to be remembered. Periodically *E* interrupts the incoming information to request *S* to recall some of the items he has been remembering. The *S* does not know when a request will occur or the serial order in which items will be requested.

Prior research indicates that at least two variables are important in determining the number of errors *S* will make in a sequence of this kind. The first is the number of items *S* is supposed to be remembering at each recall point. This number may be constant throughout a total sequence (Yntema & Mueser, 1960), or it may vary around some specified average for part of a sequence and then abruptly shift to another average (Lloyd et al., 1960). In either case recall errors increase regularly as this *average storage load* is increased.

The second variable is the number of items *S* is requested to recall at each recall point. For example, *S* may hear: pine tin polo METAL copper TREE . . . in which lower case words are items to be stored and upper case words are class names indicating which item is to be recalled. At the recall point, METAL, *S* should recall one item, tin. In another sequence: copper tin polo METAL pine TREE . . . *S* can recall two items at the first recall point. In both cases he is storing three items, i.e., his storage load is 3, but the number of items he can recall when he hears METAL is 1 in the first case and 2 in the second. Like average storage load (ASL), the number of items requested at each recall point (the *load reduction*) can vary around some average and can be abruptly changed. Reid, Lloyd, Brackett, and Hawkins (1961) reported that recall errors decreased as the average load reduction (ALR) in a sequence was increased. Both of these variables are summarizing measures of many nonindependent factors operating within the sequence. Lloyd et al. (1960) and Reid et al. (1961) have discussed some of these factors. One of them is important for the present paper, namely, that as ALR is increased there is a concomitant decrease in the number of recall points needed to request a given number of items.

---

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49 (638)-805. Experiment I in an enlarged form was a thesis submitted to the Graduate School of Washington State University in partial fulfillment of the requirements for the degree of Master of Science in psychology by the second author.

The purpose of the present work was to continue to study the effects of these two task variables and to explore the effect of a third variable, contextual constraint among items. This variable was selected because it has consistently improved recall in the classic memory span task (Miller & Selfridge, 1950; Postman & Adams, 1960; Richardson & Voss, 1960; Sharp, 1958), and because different predictions about its effects in the sequential task could be made.

Introducing contextual constraint among items in the sequential task highlights differences between this task and the memory span task. One difference is that in the sequential task several recall points would occur within a passage of, say, 20 words. These recall points would interrupt the continuity of a passage. The *S* would have to rely on "local context" rather than "general context" (cf. Postman & Adams, 1960, p. 161). Nevertheless, even isolated bits of context should help, e.g., remembering "the, boy" should still be easier than remembering "ticket, smoke" with a storage load of 6 or 7 (cf. Miller & Selfridge, 1950, p. 183). The effect of this difference should be to decrease, but not necessarily remove, the facilitative effects of contextual constraint.

Another characteristic of the sequential task requires a consideration of the *locus of constraint.* This is required since the serial order of item presentation is seldom the same as the serial order of item recall. This means that the "sense" of a passage may be located among the items as they are presented, or as they are to be recalled.

An example will illustrate what is meant by locus of constraint. Sequences are initially constructed in letter form, e.g., $a_1b_1a_2c_1d_1b_2\text{B}d_2\text{A}e_1$. . . . Next words are paired with the lower case letters and at a recall point the upper case letter alone informs *S* which words are requested. If the words paired with the letters are drawn from a contextual passage *S* hears: a-turkey b-will a-die c-on d-my b-rug B d-is A e-deep. . . . At Recall Point B the correct responses are will and rug. In this case the locus of constraint was among the items, but the context has been destroyed by the arbitrariness of the position of the recall point. Instead of predicting that recall would increase as degree of constraint increases, predicting the opposite seems more reasonable. That is, *S* should be tempted to say, "will die" at B, and hence be in error. Increasing the degree of contextual constraint in a passage when the locus of constraint is among the items should increase this tendency to recall items in a "sensible" order and, hence, to increase errors since the sensible order will usually not be correct.

In contrast to this *E* can introduce a contextual passage to correspond to the serial order of recall points. Now *S* would hear: a-die b-turkey a-on c-is d-my b-will B d-rug A e-deep. . . . The correct responses (turkey, will at B and die, on at A) correspond to language patterns, and the usual predictions for contextual constraint should hold, but to a lesser extent due to recall point interruptions.

In summary, two loci for contextual constraint are possible: among the items during presentation (stimulus order), or among the recall points (response order), and opposite predictions regarding recall can be made. In addition, these expected results may vary with both ASL and ALR. At low ASLs recall will be high regardless of context, but at high ASLs *S*s should welcome any hints, and the effects of context should become more obvious. At low ALRs the large number of recall points will interrupt passage continuity, but at higher ALRs (i.e., fewer recall points) the preserved passage continuity would permit *S* to rely on "general context" as well as "local context," and recall should improve.

EXPERIMENT I

Method

The *S*s stored and recalled letter-word pairs presented to them aurally by a magnetic tape recorder. They were instructed to remember each letter-word pair, and to recall the word when they heard the letter alone later in the sequence. Each sequence contained two ASLs (3 and 7) and four orders of approximation (0, 2, 4, and 8). Sequences were divided into eight subsequences in order to pair each load value with every order of approximation. All subsequences contained 20 items (20 letter-word pairs) and 10 recall points (10 letters alone). Two items were requested at each recall point. The individual storage loads at the recall points were 2, 3, and 4 when ASL was 3, and 6, 7, and 8 when ASL was 7. The serial order of these individual loads was varied from one subsequence to another. Six letters were used equally often and arranged in a variety of serial orders. Near the end of all but one of the subsequences it was necessary to add from 2 to 5 items in order to maintain the desired ASL. Recall points for these extra items were placed immediately after the tenth recall point in a subsequence; errors at these recall points were not included in the data analysis.

The words were obtained from three different sources. The passages for the higher orders of approximation were obtained from James E. Deese at the Johns Hopkins University. Zero order passages were made from those listed in Miller and Selfridge (1950), and words randomly selected from the most frequent English words in Thorndike and Lorge (1944). Eight different passages of 20 words each were obtained for each of the four orders of approximation. These 32 passages were presented twice to each *S*. The first presentation of a given passage occurred on Day 1 or 2 at one ASL. The second presentation of this same passage occurred on Day 3 or 4 at the other ASL.

The third variable, locus of constraint, was varied between two groups of 12 *S*s each. For the stimulus order group (Group SO) the 20 words from a passage were paired in their original serial order with the 20-item letters in a subsequence. The correct order of their recall was determined by the serial order of the 10 recall point letters in the subsequence. For the response order group (Group RO) the words in their original serial order were placed two at a time at the 10 successive recall point letters in a subsequence. Their presentation order was then determined by the serial order of the corresponding 20-item letters.

Eight sequences were recorded for each group. The serial order of the subsequences differed in each sequence. Items were recorded at a 4-sec. rate and recall points at a 3-sec. rate. Recall points were preceded by a short, 500-cps tone to aid *S* in detecting them. The *S*s, Washington State University male undergraduates, were assigned to a group on the basis of a coin toss. They were reimbursed for their services. The *E* recorded all responses. An error was defined as a failure to respond, or as a recall of a word not requested at that recall point. All *S*s responded to two sequences a day for 4 successive days.

Results

The daily means for the 12 *S*s in each group are in Table 1. An analysis of variance indicated significance for ASL, F (1, 682) = 8,943.95, $p < .01$; approximation, F (3, 682) = 23.96, $p < .01$; practice, F (3, 682) = 93.02, $p < .01$; and the Approximation × Locus of Constraint interaction, F (3, 682) = 7.95, $p < .01$. Locus of constraint was not a significant main effect, F (1, 22) = 0.35.

The significant mean differences due to ASL and to practice are clearly

TABLE 1

Mean Recall Errors for All Days for Locus of Constraint, Order of Approximations, and Average Storage Load in Exp. I

Average Storage Load	Day	Locus of Constraint	Order of Approximation							
			0		2		4		8	
			Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
3	1	SO	8.58	1.77	7.25	2.59	7.87	2.08	8.12	1.94
		RO	9.04	1.72	7.58	1.51	8.17	1.54	7.75	1.73
	2	SO	7.29	1.87	6.12	2.21	7.37	2.53	7.62	2.16
		RO	7.29	1.63	6.17	1.86	7.37	2.10	6.54	2.16
	3	SO	5.87	1.49	4.33	1.61	5.58	2.46	6.67	1.87
		RO	7.00	1.20	5.08	2.16	5.67	2.67	4.87	2.76
	4	SO	5.21	1.76	5.33	2.88	5.96	2.25	5.42	2.26
		RO	5.96	1.30	4.92	1.54	4.29	2.02	4.00	1.49
7	1	SO	17.71	0.72	16.58	0.97	17.46	1.13	17.58	1.08
		RO	17.96	0.77	16.46	1.85	17.12	1.02	17.04	1.44
	2	SO	16.96	1.14	16.00	1.41	16.62	1.62	16.87	1.57
		RO	17.37	0.65	14.83	1.74	16.33	1.45	16.54	1.26
	3	SO	16.62	1.78	15.83	1.05	16.42	1.34	16.42	0.99
		RO	17.00	0.93	15.08	1.31	15.71	1.20	16.21	1.58
	4	SO	15.71	1.62	15.29	1.67	15.67	2.19	16.29	0.93
		RO	16.37	1.25	15.04	1.82	15.00	1.56	14.71	1.51

shown in Table 1. The significant effect for order of approximation may be seen in the row means. Errors were greatest at zero order, lowest at the second order, and then increased to values between these extremes at the fourth and eighth orders. The means for both groups follow this pattern, but the slope differs for the two groups. Group RO made more errors than Group SO at the zero order, and then made fewer errors than Group SO at the higher orders. The curves for the two groups crossed once every day. This crossing of the curves was the source for the significant Locus × Approximation interaction. It also contributed to the failure to obtain a significant effect for locus of constraint. The original hypothesis specified the patterns of the curves for both groups, that is, increasing the degree of contextual constraint should increase errors in Group SO and decrease errors in Group RO. The data indicated that the effect of contextual constraint did differ in the two groups, but the patterns of the empirical curves were not the predicted ones. The hypothesis was supported, but not confirmed.

Two features of the data prohibited a confirmation of the hypothesis. First, the groups should not have differed so consistently at the zero order. Since, at the zero order, all *S*s stored and recalled the same unrelated words, it seems likely that *S*s in Group RO were initially inferior. If the groups had been equated on performance at the zero order, then differences between group means at the higher orders might have been larger. The second complicating feature of the data was the consistency with which fewest errors occurred at the second order passages. The error dip at the second order was largely responsible for the bizarre pattern of the curves for order of approximation.

The pattern of the Day 4 data, however, differed from the usual pattern. On Day 4 row means increased for Group SO and decreased for Group RO. This was the predicted result. The fact that it occurred only after 3 days of practice could mean that *S*s had been learning to detect the context among the words. Given additional practice the error trends for Day 4 might have been replicated. To test this possibility Exp. I was repeated with 6 *S*s who had already served in another sequential task study. Practice was extended to 10 days. Group RO made fewer errors than Group SO at all orders of approximation. Differences between groups were least at the zero order. The trend noted on Day 4 in Exp. I was never repeated. Instead fewest errors were still made with second order passages. Examination of the individual error curves in Exp. I and its replication showed that fewest errors occurred at the second order in 51% of them; this is double the chance value. This regularity indicated the operation of uncontrolled variables in the second order passages.

One uncontrolled relationship could have been an interaction between the ALR of 2 used in Exp. I and the second order of approximation. When ALR and order of approximation have the same value, the basic unit of the passage (digrams, trigrams, etc.) can be recalled without interruption at one recall point; if their values differ, then overlapping units, or only part of a unit, must be recalled. An experiment was designed to test the hypothesis that fewest errors would occur in those subsequences in which these variables had the same value. Second, third, and fourth order passages were inserted into subsequences with ALRs of 2, 3, or 4. The 12 *S*s served in all nine conditions. The ASL was 5, and only an RO group was tested. In the analysis of variance ALR was highly significant, but order of approximation and the ALR × Approximation interaction were not. This hypothesis was not supported.

These results seemed to eliminate the task as a source for uncontrolled factors, so the second order passages themselves were examined. An attempt was made to compare them with the passages from the other orders of approximation. The most distinguishable feature of the second order passages was the high incidence with which given words were repeated within a passage. These passages contained several strings of the sort: you are you are you. . . . In Exp. I *S*s in either group could have responded "you, are" at a recall point and *E* would have scored the responses correct without knowing which "you" or "are" the *S*s had recalled. This confusion could have happened only in rare instances at the other orders. It appeared that high word frequency within a given passage could have benefited all *S*s at second order passages.[2] It should also be noted that in strings of repeated words the chances for any pair to be contextually related was increased. That is, "local context" was undoubtedly highest at the second order.

The second order passages must have confounded the results of Exp. I. But even if the means for the second order in Table 1 are ignored the remaining differences due to contextual constraint are not impressive. Experiment II was designed to eliminate those characteristics in a sequence which seemed most likely to have destroyed context in Exp. I.

---

[2] High word frequency within the sequential task does increase recall (Lloyd, 1961a). This increase in recall occurs when the words vary in their frequency within the sequences even though they have equal English text frequency. Recall does not change if the words occur equally often in the sequences, but vary widely in their English text frequency.

EXPERIMENT II

Method

The purpose of Exp. II was to demonstrate that the effect of contextual constraint would be increased by reducing the number of recall points that occurred within a subsequence. This was done by systematically increasing ALR. The ALR in the subsequences was either 2, 4, 7, or 10 (with 20 items in each subsequence the number of recall points was 10, 5, 3, or 2, respectively). Three individual load reductions were used to obtain each ALR, e.g., with an ALR of 4 the individual reductions were 3, 4, and 5.

The ASL was always 10. The individual storage loads were always 9, 10, or 11. In those subsequences with an ALR of 10 the individual load reductions of 9, 10, or 11 always occurred with individual storage loads of the same values so that *S* was requested to recall all the words he was storing at each recall point. Except for the letters paired with the words *S*s essentially performed the same task as that employed by Miller and Selfridge (1950) for their 10-word passages. The conditions in Exp. II represented an orderly transition from the sequential task to the memory span task. It was hypothesized that as ALR increased the error curves for order of approximation would become more similar to those reported for the memory span task.

In order to enhance the effect of contextual constraint only an RO group was used. As in Exp. I the words from a passage were placed first at the recall point letters, and then paired with the corresponding item letters in their original serial order. With the high load reductions used in Exp. II most of the words *S* stored at any one time had been paired with only one letter, and, hence, *S* heard longer continuous blocks of a passage. For example, with a load reduction of 7 and a storage load of 9 *S* heard: a-turkey a-will a-die b-deep a-on a-my a-rug b-with a-is A b-snow. . . .

Eight sequences, each with eight subsequences, were constructed. Four orders of approximation (0, 2, 4, and 8) were used. Two sequences were required in order to pair all values of ALR with each order of approximation. The 12 *S*s responded to two sequences a day for 4 days, and thus experienced each condition once a day. Sixteen different word passages from each order of approximation were used. The passages were drawn from the same sources as in Exp. I. The *S*s were never presented the same passage twice.

Items were recorded at a 4-sec. rate. The duration of recall points varied from 2 to 9 sec., depending partly upon the size of the load reduction. A perfect correspondence between size of load reduction and duration of recall points was purposely avoided so that *S* could not learn to use the time interval as a clue for deciding how many words to recall. The remaining details of the experiment were the same as in Exp. I.

Results

An analysis of variance of the mean recall errors indicated significance for ALR, $F\ (3, 33) = 350.29$, $p < .01$; approximation, $F\ (3, 33) = 71.08$, $p < .01$; practice, $F\ (3, 33) = 32.22$, $p < .01$; and the Approximation × ALR interaction, $F\ (9, 99) = 43.66$, $p < .01$. Since this interaction was crucial to the hypothesis, the means for these two variables are presented in Table 2.

The column means reveal the consistent decrease in recall errors as ALR is increased. The row means decreased between the zero and second orders of approximation; beyond the second order row means changed very little. The significant Approximation × ALR interaction was largely due to differences in the rate of decrease in errors in the column means. As ALR was increased the differences between errors at the zero order and the other orders also increased. These results confirmed the hypothesis of Exp. II.

The rank order of the means in the first row in Table 2 closely resembles that for Group RO in Exp. I where the load reduction was also 2. As in Exp. I fewest errors occurred at the second order passages in 56% of the 48 individual error curves. In Exp. I it was not possible to decide which of two factors (i.e., word frequency or local context) was responsible for this

TABLE 2

MEAN RECALL ERRORS FOR ORDER OF APPROXIMATION AND AVERAGE LOAD REDUCTION IN EXP. II

Average Load Reduction	Order of Approximation							
	0		2		4		8	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
2	18.27	0.45	17.06	1.19	17.40	0.90	18.19	0.68
4	15.60	0.73	14.04	1.30	14.62	1.18	14.27	1.01
7	13.98	0.88	9.31	1.63	9.44	1.75	10.56	1.63
10	12.73	1.15	5.27	1.75	4.52	2.18	5.42	2.62

dip in errors at the second order. In Exp. II word frequency within a passage remained constant as ALR was increased, i.e., always highest at the second order, but the opportunity for *S* to respond to larger chunks of local context increased as ALR increased. It seems reasonable to argue that if word frequency is the important variable, then fewest errors should occur at second order passages for any value of ALR. Inspection of Table 2 indicates that this is not the case. Instead as ALR increased the site of fewest errors tended to shift from second order passages to fourth order passages. When ALR was 10, fewest errors occurred at the fourth order in 54% of the individual error curves, and at the second order in only 27% of the cases. This shift in the low point of the curves could indicate that *S* responded to larger chunks of contextual constraint as ALR increased.

If this interpretation is correct, then the facilitation due to local context lagged behind the opportunities for it to aid recall. That is, with an ALR and ASL of 10, *S* had an opportunity to discriminate considerable context in an eighth order passage, yet fewest errors occurred at the fourth order. With considerably higher storage loads and load reductions (as exist in the memory span task) fewest errors should eventually shift to the eighth order. The data from the memory span task indicate that such a shift does occur (cf. Miller & Selfridge, 1950, p. 182).

DISCUSSION

The main purpose of these studies was to determine the effect of contextual constraint among items in the sequential task. In Exp. I the effect was barely detectable. In Exp. II the effect was greater, but the maximum effect occurred only when the sequential task was basically changed in form to resemble the memory span task. When this occurred the present data agreed with prior data. If the means in the fourth row in Table 2 where ALR and ASL were equal are changed to percent correct responses, the curve closely resembles those for 10-word passages in Richardson and Voss (1960).

In comparing the effect of context in the two memory tasks it appears that context is very important provided *S* is not interrupted. If he is interrupted any facilitation from context prior to the interruption is likely to be lost after the interruption. This loss increases with the number of interruptions. The data do not specify the nature of these interruptions, but it may be that the request for an item plus the actual recall of the item results in a loss of other items which were also being stored but which were not requested.

Another goal of the present studies was to test further the relationship between the task variables and recall. One purpose of the prior studies using the sequential task had been to demonstrate that the effects of ASL and ALR could be reproduced in a variety of conditions. The conditions for storing items in Exp. I and II differed from previous studies. Lloyd et al. (1960) had *S*s store related words which were easily grouped together. Brackett, Reid, and Johnson (1960) used easily grouped words, but requested them in a manner that broke up the likely groupings. Lloyd (1961b) used as items unrelated words which were always paired with the same letters. In the zero order passages in Exp. I and II *S*s stored unrelated words which could be paired with any of the six letters. In all cases recall errors regularly increased with increases in ASL and decreased with increases in ALR. These relationships have now been tested over a wide range of values of the task variables.

In the earlier study of ALR (Reid et al., 1961) the better recall at higher ALRs could have been due to increased grouping of related words, or to the fewer recall points. In Exp. II at the zero order the opportunity for *S* to group words was minimal, but the number of recall points decreased as before. An analysis of variance of the zero order means (Table 2, Column 2) indicated that ALR was significant, $F\ (3, 33) = 180.51$, $p < .01$. Thus simply increasing ALR without the presence of additional factors facilitated recall. However, the introduction of other factors, such as related words, or contextual constraint, increased this facilitation.

REFERENCES

BRACKETT, H. R., REID, L. S., & JOHNSON, R. B. Short-term retention in sequential task performance: III. Variable conditions determining the influence of average load reduction upon retention. Technical Report No. 537-60-3, August 1960, Office of The Surgeon General.

LLOYD, K. E. Short-term retention as a function of item frequency. *Amer. Psychologist*, 1961, **16**, 465. (Abstract) (a)

LLOYD, K. E. Supplementary report: Short-term retention as a function of average storage load. *J. exp. Psychol.*, 1961, **62**, 632. (b)

LLOYD, K. E., REID, L. S., & FEALLOCK, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–186.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

POSTMAN, L., & ADAMS, P. A. Studies in incidental learning: VIII. The effects of contextual determination. *J. exp. Psychol.*, 1960, **59**, 153–164.

REID, L. S., LLOYD, K. E., BRACKETT, H. R., & HAWKINS, W. F. Short-term retention as a function of average storage load and average load reduction. *J. exp. Psychol.*, 1961, **62**, 518–522.

RICHARDSON, P., & VOSS, J. F. Replication report: Verbal context and the recall of meaningful material. *J. exp. Psychol.*, 1960, **60**, 417–418.

SHARP, H. C. Effect of contextual constraint upon recall of verbal passages. *Amer. J. Psychol.*, 1958, **61**, 568–572.

SHEPARD, R. N., & TEGHTSOONIAN, M. Retention of information under conditions approaching a steady state. *J. exp. Psychol.*, 1961, **62**, 302–309.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, Bureau of Publications, 1944.

YNTEMA, D. B., & MUESER, G. E. Remembering the present states of a number of variables. *J. exp. Psychol.*, 1960, **60**, 18–22.

(Received June 26, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 468–473

PERCEPTION OF LINEAR DOT PATTERNS[1]

E. T. KLEMMER[2]

International Business Machines Corporation Research Center, Yorktown Heights, New York

Linear dot patterns were presented tachistoscopically by flashing randomly selected patterns on uniformly spaced light bulbs. The total visual angle was varied from 2.5° to 160° using 21 bulbs and, in a separate experiment, number of bulbs was varied from 11 to 35, using a fixed 1° spacing between bulbs. Perception was determined by a poststimulus cue requiring only a 1-bit response. Average accuracy of report was surprisingly insensitive to visual angle. The center bulbs were seen better than the end bulbs at all visual angles. When the spacing of bulbs was fixed at 1° and the number of bulbs increased from 11 to 35 the percent-error increased linearly with near perfect performance for 11 bulbs (poststimulus cue method). Comparison with previously reported data on the perception of linear dot patterns indicates that the earlier findings may have been due to memory rather than perception.

One of the questions blocking the development of a comprehensive theory of pattern perception concerns how much of the stimulus field to consider as influencing any given momentary perception. Stated more specifically: how should information bearing elements be weighted as a function of their visual angle from the point of fixation? The statement of the question does not presuppose a simple answer. The present study provides a partial answer to this question for patterns formed by dots placed along a single perimeter, varying in number and over a wide range of visual angles.

Linear dot patterns have been studied previously by Pollack and Klemmer (1954) and Harcum (1957). These studies used eight and nine binary positions, respectively, with short exposures. Harcum varied orientation of the pattern but neither study varied the overall visual angle or the number of binary positions. It also turns out that requiring *S* to reproduce the entire pattern in these earlier studies makes the results much more a function of memory of binary sequences than perception of binary patterns.

Several studies, e.g., French (1953), have concerned the discrimination or detection of two-dimensional dot patterns. This work has been largely inspired by PPI radar displays and, therefore, has not varied size over large visual angles and has emphasized detection rather than perception. Eventual perceptual theories should include both types of behavior, but it is still premature to attempt such integration. The present experimental study describes perception in a situation in which all of the presented information is equally pertinent and this information is uniformly spread, in one dimension, over a wide range of visual angles. Under these rules then, what are the invariant relations in linear dot pattern perception?

---

[1] G. R. Lockhead and J. A. Tirrell prepared the stimulus materials, conducted the experimental sessions, and assisted in the data analysis. H. Serenson programed the IBM 704 computer to reduce the raw data. Their contributions to this study are gratefully acknowledged. An earlier version of this paper appeared as a technical report (Klemmer, 1961).

[2] Now at the Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey.

Three somewhat overlapping experiments were conducted. The first, varied overall visual angle with a fixed number of bulbs. The second, gathered more extensive data on three of these conditions to make possible pattern and positional analysis. The third experiment held spacing between bulbs fixed and varied number of bulbs and overall visual angle. All experiments used the same apparatus.

General Method

Apparatus and common procedure.—The viewing field consisted of a visual perimeter 10 ft. in radius. The perimeter provided a field uniform in 180° of azimuth with a near turret which limited the vertical view of the perimeter to 5°. Up to 35 small incandescent bulbs could be clipped on the perimeter at any position subject only to minimum spacing of $\frac{1}{4}$ in. (7′ of arc). The background colors were white above the perimeter line and brown below with brightness of .2 and .1 mL., respectively. The bulbs themselves were visible at all times but no other index marks were present except for a small projection over the center bulb which served as a fixation point. The *S* fixated on the center bulb and then triggered a .04-sec. flash of a random pattern of bulbs selected by an IBM card reader. One-half second after the flash, without further action by *S*, one of the bulbs came on dimly. This bulb, chosen at random without regard to the previous pattern, served as the poststimulus cue. Another $\frac{1}{2}$ sec. later the stimulus field was restored to the flashed pattern for the "restored" condition, with the exception of the cue light which remained dimly illuminated regardless of its flashed state. For the "nonrestored" condition only the poststimulus cue light appeared after the stimulus. The *S* reported only whether the cued bulb was lighted or dark in the original flashed pattern. A second button, operated by *S* after his response, restored the cue bulb to its flashed condition, giving him immediate knowledge of results. A third button cycled the card reader ahead to the next stimulus. All response recording was by paper and pencil by *E* in the room with *S*. Usually *S*s and *E*s alternated roles in the middle of the daily experimental session which took about 20 min. for each half. Staff members served as *E*s in enough sessions to ascertain that no bias was present in the usual procedure.

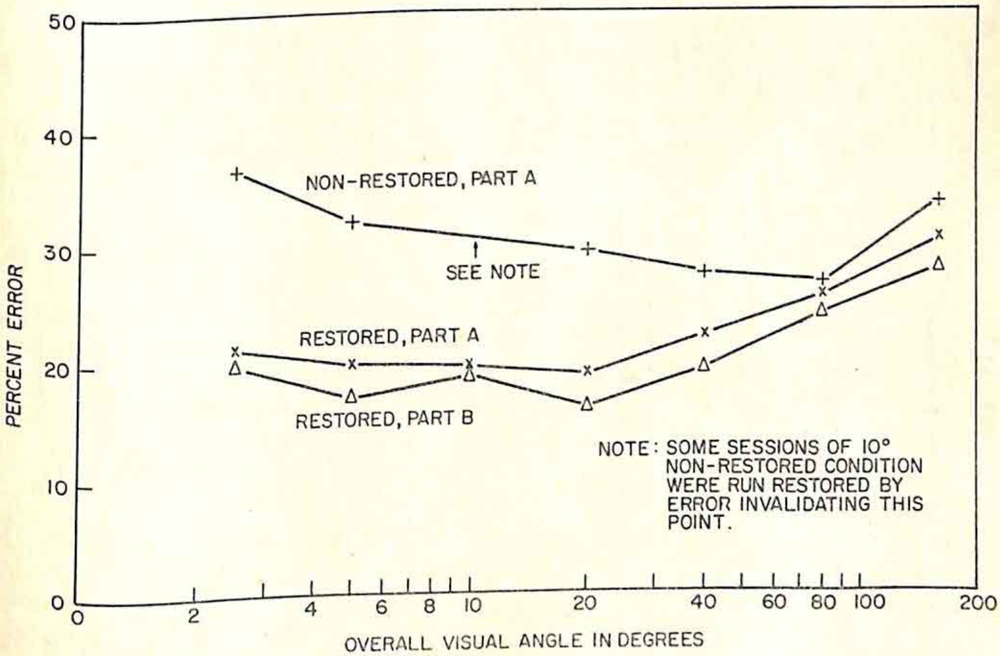


Fig. 1. Error percentage as a function of overall visual angle for linear dot patterns formed on 21 bulbs. (Combined data from 4 *S*s who reported only one bulb position as designated by a poststimulus cue. Each point represents 1,008 or 2,016 observations.)

EXPERIMENT I: EFFECT OF OVERALL VISUAL ANGLE

Method

Subjects.—Four college students served as paid *S*s in Part A and 2 of these served with 2 new *S*s in Part B. Each had at least two sessions of training, demonstrating the range of conditions to be used before starting the experimental sessions.

Procedure.—Each of the *S*s was run on the same balanced design as described below using the same random order of patterns. The presentation sequence was divided into sessions of 126 stimuli with bulb locations fixed during the session.

The number of bulbs was fixed at 21 throughout Exp. I. Part A consisted of 28 sessions with the poststimulus restoration of the pattern alternated each session between "restored" and "nonrestored." The overall visual angle was changed every other session in logarithmic steps from 2.5° to 160° and then from 160° to 2.5°.

Part B also involved 28 sessions but all sessions were run under the restored condition and the overall angle was changed after each session in a descending-ascending design in the same logarithmic steps from 160° to 2.5°, 2.5° to 160° and repeat. Thus for each of the seven overall visual angles each *S* saw 252 stimuli under each condition in Part A and in Part B each *S* saw 504 stimuli under the restored condition.

The choice of bulb to be cued was random within the restriction that each of the 21 bulbs was cued six times each session.

During the first 28 sessions (Part A), in addition to the binary response regarding the cued bulb, *S*s were asked to estimate the confidence of their judgment on a five-point scale; 5 for certainty, 1 for a complete guess. Confidence judgments were not used in any other part of this study.

Results

The average percentage of error for each overall visual angle is given in Fig. 1 for both the restored and nonrestored conditions. The first and second 28 sessions are plotted separately for the restored conditions to allow comparison with the nonrestored condition which was run only during the first 28 days.

As would be expected, the restored condition is better than the nonrestored at all angles but the difference is larger for the smaller angles. All angles show an improvement with practice on the restored condition

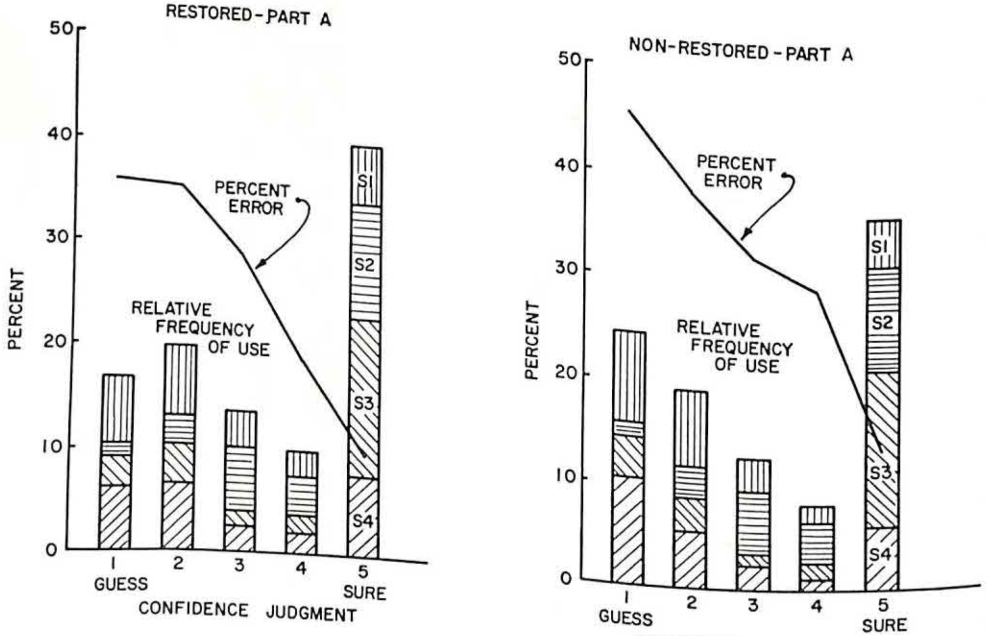


FIG. 2. Relative frequency of use of confidence judgments and relation to actual error rate. (Each *S*'s contribution to the frequency-of-use bars is shown separately. The line relating error rate to confidence judgment is the result of combining the raw data from 4 *S*s.)

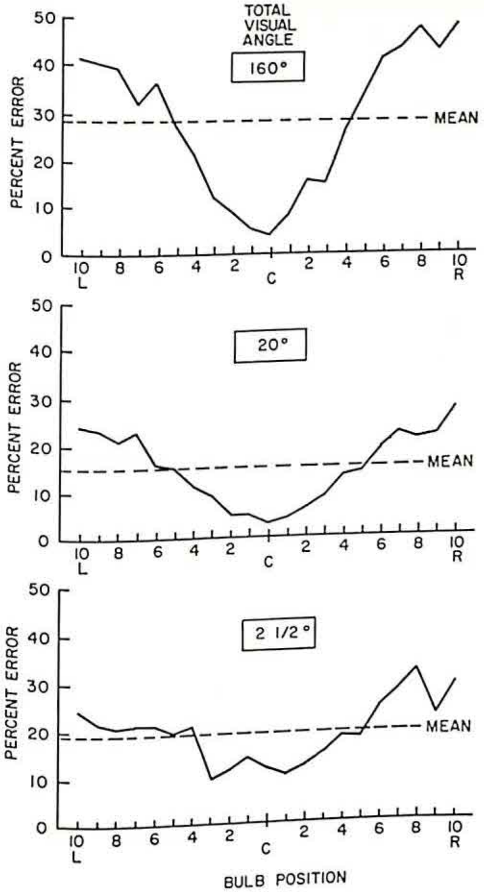


FIG. 3. Mean error rate as a function of position of cued bulb. (Data is the average of 4 *Ss*, each point representing 288 observations.)

between Parts A and B but this difference is only about 2%. Most interesting, however, is the relatively small effect of overall visual angle. Its variation in a 64 to 1 range produces a variation in error percentage of less than 2 to 1.

The relation between confidence judgments and actual percentage correct is shown in Fig. 2, along with the relative frequency of use of each confidence level. The data from the restored and nonrestored conditions are shown separately. The contribution of each of the 4 *Ss* is shown by shadings of the frequency-of-use bars in Fig. 2.

Confidence judgments and error percentage are negatively related, but the lowest confidence level shows better than chance performance and the highest level (instructed to denote certainty) still shows an average error of 10 to 14%. Averaging over all *Ss* the frequency-of-use bars in Fig. 2 show the certainty judgment (5) to be the most frequently used and the adjacent category (4), the least frequent.

EXPERIMENT II: PATTERN AND POSITION EFFECTS

Method

Subjects.—The same 4 *Ss* who served in Part B of Exp. I were employed in Exp. II.

Procedure.—The procedure was the same as Exp. I, Part B, except that only three total angles (160°, 20°, 2.5°) were used in order to build up sufficient data to allow an analysis of performance for each of the 21 bulbs and for patterns surrounding the cued

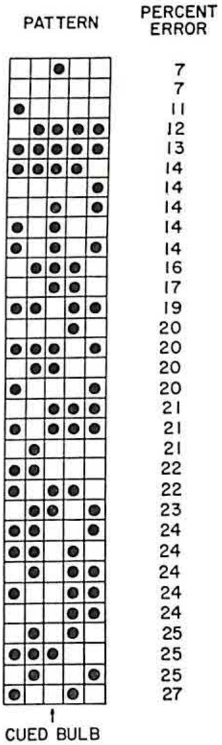


FIG. 4. Order of difficulty of the 5-bit pattern surrounding the cued position. (Error percentages are based on the combined data of 6 *Ss* totaling about 1,000 observations per pattern.)

bulb. Thirty-six sessions were given with each *S* proceeding through an ascending-descending design (2.5°, 20°, 160°, 160°, 30°, 2.5°) repeated three times. The total visual angle was changed every second session, two sessions being given each day, 126 stimuli per session.

Results

The average percentage correct score for all *S*s and each bulb position and each total visual angle is shown in Fig. 3. The center bulbs are always seen better than the ends with no indication of a drop in error rate at the end bulbs (end effect). The three curves of Fig. 3 make it clear that the poorer performance of 2.5° than on 20° is due to increased errors on the middle bulbs while poorer performance at 160° is due to more errors on the end bulbs.

For analyzing the influence of the particular pattern it is clearly impossible to work with more than two million possible patterns of the 21 bulbs. It is possible to analyze the effect of patterns formed by the cued bulb and the two bulbs on each side of it. The five-bulb patterns thus formed represent 32 different alternatives. In order to get reasonably stable estimates of the relative performance on each pattern, the results from all *S*s have been combined from all sessions of Exp. II and the "restored" sessions of Exp. I. The resulting order of difficulty of the 32 patterns is shown in Fig. 4. Note that there is little differentiation among the more difficult patterns, but the easiest two patterns have less than one third the error rate of the more difficult patterns. The 11 easiest patterns all include symmetrical patterns in the center three bulbs (on-on-on, off-off-off, or off-on-off). The one remaining symmetrical 3-bit pattern (on-off-on) occurs only in patterns of average or poor performance. A word of caution is in order about Fig. 4. In order to be reasonably certain that there is a real difference between patterns, the difference in error rate must be at least 4%. This figure is based on the 90% confidence range of the differences in performance between mirror images of the nonsymmetrical patterns.

EXPERIMENT III: FIXED SPACING

Method

Subjects.—The 4 *S*s of Exp. II served in Exp. III.

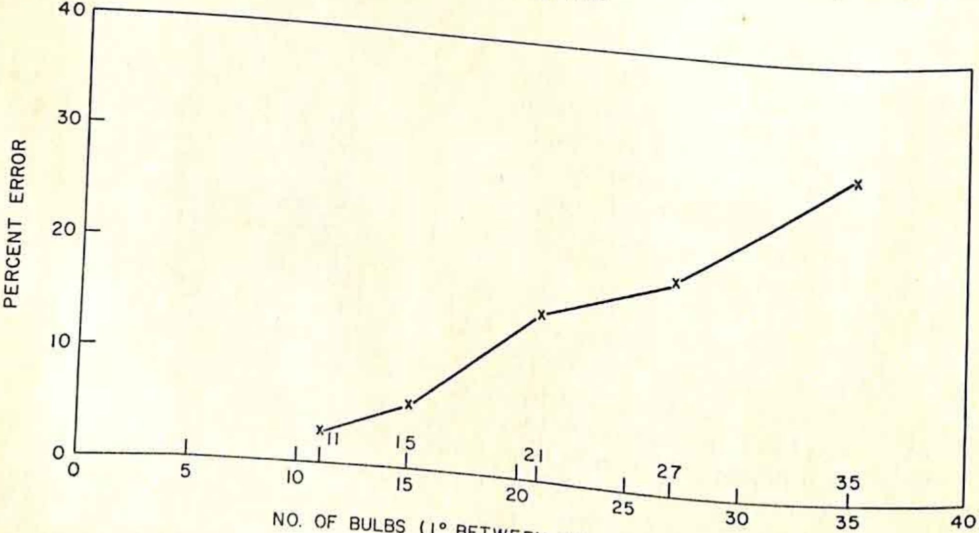


FIG. 5. Error rate as a function of number of bulbs when adjacent bulbs are uniformly spaced at 1°. (Combined data from 4 *S*s.)

Procedure.—Each *S* was run on the same balanced design as described below and the same random order of patterns. The presentation sequence was divided into sessions of about 120 stimuli with bulb locations fixed during the session and changed between sessions.

The spacing between bulbs was fixed at 1° of visual angle for all sessions. The total number of bulbs was changed in the following sequence over 10 sessions: 11, 15, 21, 27, 35, 35, 27, 21, 15, 11. The unused bulbs were covered so that the active bulbs were clearly bracketed. All trials were run under the restored condition; that is, the pattern, with the exception of the cued bulb, was reilluminated before the response.

Results

The average performance in terms of percent-error is shown in Fig. 5 as a function of the number of bulbs. There is very little error with 11 bulbs and a linear increase in percent-error with number of bulbs from 11 to 35 bulbs. (This function must asymptotically approach 50% error and, therefore, cannot be linear much past 35 bulbs.)

DISCUSSION

Experiments I and II demonstrate the insensitivity of the tachistoscopic perception of linear dot patterns to large changes in overall visual angle when the number of dot positions is held constant. This is in sharp distinction to the recent results of Shinkman (1961), who showed a large loss in the perception of Arabic numerals when visual angle was changed only from 1½° to 3°. A similar contrasting finding has been reported by Osborne, Quastler, and Tweedell (1955) in comparing tachistoscopic perception of letters with scales and dials. The letters and numbers used in both of these studies were not so small as to be below the acuity threshold for the retinal locations involved so the explanation must be in terms of an interaction or some sort of informational acuity measure. Further theoretical and experimental work is needed here.

The influence of position in the visual field as shown in Fig. 2 provides an interesting comparison with similar functions given by Harcum (1957) and Pollack and Klemmer (1954). Both of the previous studies reported lowest errors at the left end, next lowest at the right end, and an inverted U relation between the ends. In contrast, the present study shows lowest errors at the middle with an erect U relation over the field. It is not difficult to explain this difference. The previous studies used only nine and eight dot positions and Fig. 5 shows that these patterns would be perceived perfectly as demonstrated by the poststimulus cue technique. The requirement to reproduce the entire pattern in the earlier studies apparently invoked forgetting and/or interference errors and thus the positional effects which are typical of memory experiments. They do not include errors due to perceptual overload which degenerate the outside elements of the field most severely.

REFERENCES

FRENCH, R. S. The discrimination of dot-patterns as a function of number and average separation of dots. *J. exp. Psychol.*, 1953, **46**, 1–9.

HARCUM, E. R. Three inferred factors in the visual recognition of binary targets. In J. W. Wuffeck & J. H. Taylor (Eds.), *Form discrimination as related to military problems.* (NAS-NRC Publ. No. 561) Washington, D. C.: National Academy of Sciences—National Research Council, 1957. Pp. 32–37.

KLEMMER, E. T. The perception of linear dot-patterns. Report No. RC-383, 1961, International Business Machines Corporation Research Center, Yorktown Heights, New York.

OSBORNE, J. W., QUASTLER, H., & TWEEDELL, K. S. Human performance in information transmission. Part III. Report No. R-68, 1955, University of Illinois, Control Systems Laboratory.

POLLACK, I., & KLEMMER, E. T. The assimilation of visual information from linear dot-patterns. *USAF Operat. Appl. Lab. tech. Rep.*, 1954, No. 54-16.

SHINKMAN, P. G. Supplementary report: Perception of stimuli of varying dimensionality. *J. exp. Psychol.*, 1961, **62**, 626–627.

(Received July 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 474–477

SODIUM AMYTAL AND PARTIALLY REINFORCED RUNWAY PERFORMANCE[1]

ALLAN R. WAGNER

Yale University

Acquisition performance of rats in a simple runway was investigated as a function of percentage reinforced trials (100% vs. 50%) and injection solution received prior to daily training (sodium amytal vs. isotonic saline) in a 2 × 2 factorial design. The results obtained from 4 groups of 12 Ss showed 50% reinforcement to lead to faster final speeds than 100% reinforcement with saline control injections; whereas this order was reversed under training with sodium amytal. The results were predicted from a frustration theory interpretation of random partial reinforcement and the assumption that anticipatory frustration is particularly susceptible to the depressant action of sodium amytal.

A relatively consistent finding in appetitional runway studies (e.g., Goodrich, 1959; Haggard, 1959; Wagner, 1961; Weinstock, 1958) is that rats will attain faster running speeds over all but the last segments of an alley, if rewards are received on only a random 50% of the trials than if they are received on every trial. Theoretical interpretations of this finding have frequently appealed to motivational differences associated with the anticipatory goal responses conditioned under the two reinforcement schedules. For example, Wagner (1961) has assumed that while *anticipatory reward*, or incentive motivation, is greater under continuous reinforcement, a motivational advantage remains under partial reinforcement, due to the added drive effects of *anticipatory frustration*.

This position, in conjunction with the familiar assumption (e.g., Wilson, Weiss, & Amsel, 1955) that partially reinforced *S*s learn to approach in the presence of anticipatory frustration as well as anticipatory reward, allows the deduction that partially reinforced *S*s will approach faster than continuously reinforced *S*s. The position also suggests, however, that if the drive effects of anticipatory frustration were removed, partially reinforced *S*s would respond slower than continuously reinforced *S*s.

In a recent report Barry, Wagner, and Miller (1962) advanced the argument that anticipatory frustration, like fear, may be particularly susceptible to the depressant action of alcohol and sodium amytal. If this argument is sound, a manipulation is available for reducing anticipatory frustration, and consequently for testing the above prediction. That is, while partially reinforced *S*s should run faster, asymptotically, than continuously reinforced *S*s following saline control injections, they should run slower following similar injections of either alcohol or sodium amytal. The following experiment was designed to evaluate this prediction in the case of the latter drug.

METHOD

Subjects.—The *S*s were 48 experimentally naive, male albino rats of Sprague-Dawley strain, purchased from the Charles River Breeding Laboratories. Their ages ranged from 100 to 110 days at the beginning of experimental training.

---

[1] This research was supported in part by National Science Foundation Grant G-13080.

Apparatus.—The apparatus was an enclosed, straight alley with gray, pine, side and end walls and hardware cloth ceiling and floor. The alley was $3\frac{1}{2}$ in. wide and 4 in. high throughout, and divided by guillotine doors into three sections, a 15-in. start box, a 39-in. runway, and an 18-in. goal box. Two measures of locomotor performance were provided. Starting time was recorded from the raising of the start door until the interruption of a light beam 6 in. distant. Running time was recorded from the breaking of the first beam until the interruption of a second beam located 12 in. from the end wall of the goal box. A metal food cup, $1\frac{1}{4}$ in. in diameter, was permanently affixed to the end wall.

The apparatus was located in a quiet, dimly illuminated room maintained at approximately 72° F. The *S*s were also housed, in individual cages, in this same room throughout the experiment.

Procedure.—Eight days prior to acquisition training *S*s were placed on a 24-hr. deprivation schedule. Each *S* received 12 gm. Purina lab chow checkers each day, 1 hr. after its daily training. Water was freely available. During the course of these 8 days *S*s were habituated to the experimental situation by systematic handling, exposure to the reward pellets (.045 gm. each, purchased from the P. J. Noyes Company) scattered on a raised platform, exposure to the runway in groups of 3, and finally, a direct placement in the goal box with seven reward pellets in the food cup.

This preliminary training was followed by a total of 64 acquisition trials in the alley, 2 trials on Days 1 and 2, and 4 trials on subsequent days. On each day of acquisition *S* was given an intraperitoneal injection of drug or saline and replaced in its home cage. Ten minutes later *S* was introduced into the start box. After approximately 5 sec. the start-box door was raised, allowing *S* to traverse the runway. When *S* entered the goal box, the goal-box door was lowered to prevent retracing. The *S* was removed from the goal box immediately after eating, or in the case of nonreward trials, after 20 sec., and returned to its home cage. The *S*s were run in squads of 6, each trial in rotation. This produced an average intertrial interval of 4–5 min., and allowed all daily trials to be run within the 10–30 min. period following injection.

Twelve *S*s were assigned to each of four experimental groups, distinguished by the combination of reinforcement schedule and injection solution received. Two groups received injections of 20 mg/kg sodium amytal (2 ml/kg of a 10 mg/ml solution of isotonic saline) while the other two received equal volumes of isotonic saline. These dosages were the same as those used by Barry, Wagner, and Miller (1962). Within each injection condition one group received 100% reinforcement in the alley and the other 50%. The schedule of rewards for the 50% groups over the 64 trials was obtained by repeating the following sequence: + − + − − + + − − − + + + − − +. All *S*s received seven pellets on rewarded trials.

The experiment was conducted in two replications, 3 mo. apart, by different *E*s each running 6 of the *S*s in each of the experimental groups.[2]

RESULTS AND DISCUSSION

For purposes of analysis each *S*'s individual starting and running times were converted to speeds in ft/sec. Since the two response measures yielded essentially identical results, only one measure will be reported here. Running speed was chosen for presentation since it was less likely to be influenced by inter- or intra-*E* variability in placement of *S*s in the start box or in operation of the start door.

Figure 1 presents, for the two replications combined, the mean running speeds of the four experimental groups over successive blocks of eight acquisition trials. As may be observed, the two sodium amytal groups were not only slower than the two saline control groups, but the final order of the 50% and 100% groups was reversed in the two injection conditions. Whereas the 50% saline group ran faster than its corresponding 100% group, the 50% sodium amytal group ran slower than its corresponding 100% group.

Analysis of variance appropriate for a factorial design was performed on the mean speeds over the last 16 trials. The results of this analysis support the reliability of both the

---

[2] The first replication was run by Merle F. Goldstein, the second by Mary W. Miller.

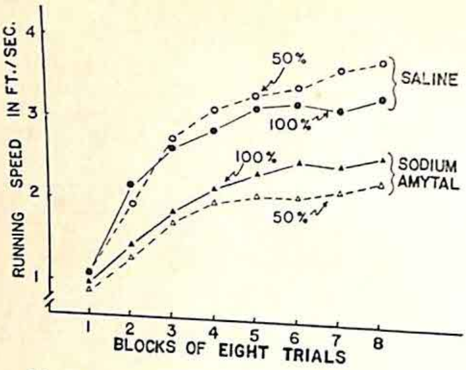


FIG. 1. Mean running speeds in blocks of eight trials for the four experimental groups.

overall depressing effect of sodium amytal, $F\ (1, 40) = 67.81$, $p < .001$, and the interaction of this effect with the reinforcement schedule, $F\ (1, 40) = 7.64$, $p < .01$. Subsequent t tests relevant to this interaction, employing the within-cells MS as the estimate of error variance, revealed 50% reinforcement to produce significantly faster speeds than 100% reinforcement under saline, $t\ (40) = 2.26$, $p < .025$, one-tailed, but significantly slower speeds than 100% reinforcement under sodium amytal, $t\ (40) = 1.70$, $p < .05$, one-tailed.

A comparison of the results from the two replications further attests to the stability of these findings. The general picture presented in Fig. 1 was obtained in each replication, with no statistically significant difference between the two in overall drug effect, $F\ (1, 40) = 2.75$, $p > .10$, or in the Solution × Reinforcement Schedule interaction ($F < 1$), although the absolute speeds obtained by one E were consistently higher than those obtained by the other E, $F\ (1, 40) = 27.75$, $p < .001$.

The results are clearly in accord with the prediction made. The general reduction in running speeds produced by sodium amytal is consistent with the drug's classification (e.g., Goodman & Gilman, 1955, Ch. 9) as a central nervous system depressant. The exaggeration of this effect under partial, as compared to continuous, reinforcement training, however, requires a more detailed theoretical interpretation. The dual assertion, that partially reinforced approaching is in part mediated and energized by anticipatory frustration, and that sodium amytal selectively reduces this emotional response, provides such an interpretation. The observation that partially reinforced *S*s ran slower than continuously reinforced *S*s, when the effects of anticipatory frustration had presumably been reduced via sodium amytal, is also in agreement with the assumption that anticipatory reward is lower under partial than under continuous reinforcement.

It should not be overlooked that random partial reinforcement has been specifically treated (e.g., Wagner, 1961) as a special case in which *S*s are reinforced for continuing to approach in the presence of anticipatory frustration. Without such special training, anticipatory frustration has been assumed (e.g., Wagner, 1961) to elicit avoidance responses. Consequently, a reduction in anticipatory frustration should, by reducing avoidance behavior, more typically produce an increase in previously frustrated responses. Barry, Wagner, and Miller (1962), for example, observed that administration of sodium amytal was followed by an increase in performance previously reduced by experimental extinction.

Considering both the present results and those obtained by Barry, Wagner, and Miller within the context of existing frustration theory, the single assumption, that sodium amytal reduces anticipatory frustration, is able to account for empirically quite diverse effects of the drug. This fact is, at least, encouraging to the sodium amytal assumption as well as to the several tenets of frustration theory involved.

REFERENCES

BARRY, H., III, WAGNER, A. R., & MILLER, N. E. Effects of alcohol and amobarbital on performance inhibited by experimental extinction. *J. comp. physiol. Psychol.*, 1962, **55**, 464–468.

GOODMAN, L. S., & GILMAN, A. *The pharmacological basis of therapeutics.* (2nd ed.) New York: Macmillan, 1955.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

HAGGARD, D. F. Acquisition of a simple running response as a function of partial and continuous schedules of reinforcement. *Psychol. Rec.*, 1959, **9**, 11–18.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hr. intertrial interval. *J. exp. Psychol.*, 1958, **56**, 151–159.

WILSON, W., WEISS, I. J., & AMSEL, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–60.

(Received July 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 478–484

S-R RELATIONSHIPS AND REACTION TIMES TO NEW VERSUS REPEATED SIGNALS IN A SERIAL TASK[1]

PAUL BERTELSON

Université Libre de Bruxelles

Previous experiments showed that, in a serial responding task, RTs to repeated signals are longer than RTs to new signals. The influence of S-R relationships on this phenomenon was examined. In Experiment I, 9 Ss gave 550 responses under each of 3 S-R conditions in a 2-choice task with a spatial display. RTs to both repeated and new signals were significantly (.01) affected by the conditions, but covariance analysis shows that the effect is significantly (.05) larger on RTs to new signals. These results are fully confirmed by Experiment II where 33 Ss, after prolonged practice on a 4-choice task with a numerical display, under a straightforward S-R pairing, did 4 short runs of 50 responses under the same pairing and then under a less compatible one.

In a previous paper (Bertelson, 1961), it was shown that in a serial self-paced responding task, reaction times (RTs) to repeated signals are shorter than RTs to signals which are "new," i.e., different from the immediately preceding one. It was proposed that this phenomenon be called "repetition effect." The effect was shown to be at least partly transitory, being greatly reduced when a .5-sec. time lag is introduced between the end of the response and the appearance of the next signal.

The existence of the repetition effect suggests either (*a*) that different mechanisms are involved in reactions to repeated signals and in reactions to new signals, or (*b*) that the same mechanisms are involved but work faster in the case of repetitions, due to some sort of facilitative aftereffect. One way of deciding between these alternatives is to consider the influence on the RTs to the two categories of signals, of factors which are known to affect overall RT. Such factors are: number of alternative signals, signal probability, and signal-response relationships. The main purpose of the present experiments was to consider the effect of the latter variable.

The choice of an S-R relationship involves two operations: the selection of sets of stimuli and responses, and the matching of sets of points in the stimulus space with those in the response space (Deininger & Fitts, 1955, p. 319). Fitts has shown that one can affect "S-R compatibility" either by modifying the matching of elements of two constant sets (Fitts & Deininger, 1954) or by changing the set of stimuli which is coupled with a constant set of responses (Fitts & Seeger, 1953). In Exp. I, both types of variation were used: two lights lying on the same horizontal axis were matched with two similarly disposed keys, either in the straightforward way or in the inverted one, and these two conditions were compared with one where the same set of keys was coupled with two lights lying on a vertical axis. In Exp. II only the matching of elements was varied.

A further way of affecting mean RT was tried out in Exp. I. It consists of having the *S* work with one S-R condition immediately at the end of a period of work with the opposite pairing. In this way one could hope to obtain an interference effect on RT, which would amount to "short-term incompatibility."

---

[1] This research was partly supported by a Crédit aux Chercheurs from the Belgian Fonds National de la Recherche Scientifique.

EXPERIMENT I

The task used in this experiment was nearly the same as in the previous experiments (Bertelson, 1961). The display consisted of two spatially separated lamps and the controls of two keys. The main difference was the fact that the relative positions of the two lamps could be varied so as to change S-R compatibility. Another difference was the introduction of an automatic recording of RTs to new and to repeated signals.

Method

Apparatus.—The display consisted of two separate pairs of neon lamps, inserted into a vertical panel. Only one pair was used in each condition. The lamps of the *horizontal* pair were 5 cm. apart horizontally, the lamps of the *vertical* pair 5 cm. apart vertically. The whole set of lamps formed a Y shaped pattern. Two bakelite keys, resting on microswitches, were placed in a near-vertical plane, 15 cm. apart, behind an inclined board. The *S* rested his hands on the upper edge of the board, his finger tips poised over the keys (see Fig. 1).

Three different S-R conditions were used: (*a*) *Direct* (D): horizontal pair of lamps, the correct response being left key for left lamp and right key for right lamp; (*b*) *Crossed* (C): horizontal pair of lamps, the correct response being left key for right lamp and right key for left lamp; (*c*) *Perpendicular* (P): vertical couple of lamps, the two possible combinations (high-right/low-left and high-left/low-right) being given to alternate *S*s.

When either key was pressed, the stimulus light went off. The next signal appeared .05 sec. after the key was released. The task was noncorrective: incorrect responses had the same effect on the display as correct ones, but they activated an error counter.

The sequence of signals was controlled by a programing unit, based on the use of two uniselectors, which has been described previously (Bertelson & Davidson, 1959). This programmer generates a series which is not repeated before thousands of steps. In the present experiment, both signals were equiprobable, i.e., the series was random as far as the principle of the programmer permitted (Bertelson, 1961, p. 91).

The apparatus stopped automatically after 50 responses. The RTs to repeated signals and to new signals were cumulated on two separate chronoscopes. Electromagnetic counters counted signals pertaining to both categories, so that mean RT in each category could be obtained for each run. The RT to the first signal in each run, which was neither new nor repeated, is excluded automatically from the recording.

Subjects.—Nine male adults, aged 21–27, all postmen learning machine sorting, served as *S*s. They were tested during normal school hours. All had already served in another experiment with the same apparatus, about 6 mo. before, but with a different display (numerical indicator), in which they had made 2,000 responses each.

Procedure.—Each *S* participated in three sessions. On each session he gave 11 runs of

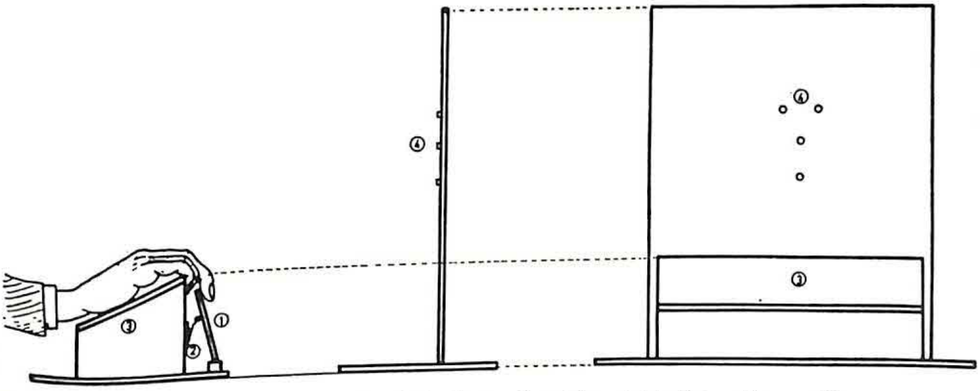


FIG. 1. View of the task: left, from the side; right, from *S*'s position. (1 = key, 2 = microswitch, 3 = resting board, and 4 = display lights.)

50 responses on one of the three S-R conditions, the order of conditions from session to session being organized in Latin squares. The first run was considered as practice and discarded for the analysis. Between 2 runs, there was an interruption of about 45 sec. during which *S* was told the total RT and the number of errors for the previous run. The *S*s were asked to go as fast as possible and to avoid errors. Total results per session were recorded in a table, which was displayed for inspection by all *S*s.

Immediately after the 11 runs which constituted the main task for the session ("main condition runs"), *S*s were asked to do 2 runs with the opposite S-R relationship, i.e., C after D; D after C; and after one P combination, the other P combination ("interference runs").

Results

Main condition runs.—For each *S*, the mean RTs to new and to repeated signals were obtained for each session. The averages over all 9 *S*s for each condition are given in Table 1. If we measure the repetition effect by the difference between RTs to new vs. repeated signals, it can be seen from the third column that it is larger in both the C and P conditions than in the D condition. Another way of looking at the phenomenon is to consider the three bottom lines of the table: much larger changes are induced by modifications of S-R relationships in RT to new signals than in RT to repeated signals.

TABLE 1

RTs IN THE MAIN CONDITION RUNS (IN MSEC.): EXP. I

Cond.	Signals		Difference New − Repeated
	New	Repeated	
Direct (D)	377	307	70
Crossed (C)	443	333	110
Perpendicular (P)	410	320	90
Differences between conditions			
C − D	66	26	40
P − D	33	13	20
C − P	33	13	20

Analysis of covariance was applied to RTs to new signals with RTs to repeated signals serving as predictor. The results are given in Table 2. A significant condition effect is observed. This means that the whole of the effect of conditions on RT to new signals cannot be predicted on the basis of

TABLE 2

ANALYSIS OF VARIANCE AND OF COVARIANCE OF MAIN CONDITION RUNS: EXP. I

Source	Variance Analyses				Covariance Analysis of y			
	df	*MS*			*df*	Adjusted *SS*	*MS*	*F*
		x^2	xy	y^2				
*S*s								
Between Latin squares	2	1.303	1.984	3.404	2	2.897	1.448	
Within Latin squares	6	1.860	−863	1.579	6	7.220	1.203	
Conditions (C)	2	1.432**	3.565	9.636[a]**	2	6.067	3.033	11.1[a]*
Order (O)	2	2.137**	514	1.718	2	3.294	1.647	10.6[b]*
C × Sq. (b)	4	98	412	312*	4	1.096	274	4.7*
O × Sq. (b)	4	138	− 43	660	4	624	156	2.7
Residual	10	119	−114	54	9	527	58.6	

Note.—y = RT to new signals; x = RT to repeated signals.
[a] Tested vs. C × Sq. (b).
[b] Tested vs. O × Sq. (b).
* $p < .05$.
** $p < .01$.

TABLE 3

RT IN THE TWO LAST MAIN CONDITION RUNS OF EACH SESSION AND IN THE TWO INTERFERENCE RUNS UNDER THE SAME CONDITION: EXP. I

Cond.	Runs	Signals		Difference New − Repeated
		New	Repeated	
Direct	Main condition	373	304	69
	Interference	389	304	85
	Difference	16*	0	16
Crossed	Main condition	437	322	115
	Interference	453	335	118
	Difference	16	13	3
Perpendicular	Main condition	407	317	90
	Interference	436	319	117
	Difference	29*	2	27

* Difference between main condition runs and interference runs significant at $p < .05$.

the effect on RT to repeated signals. The latter effect is significant, as appears in the analysis of variance which is also given in Table 2.

Interference runs.—The RTs for the two interference runs of each session were compared with those for the last two main condition runs under the same condition. This means that for Cond. D and C the comparison is made with runs performed on another session, while for the P condition, it is made with the two immediately preceding runs. The results which are given in Table 3 show that the repetition effect is always larger in the interference runs. Analysis of covariance was applied separately to the data for each condition. A significant interference effect was found for the P condition, $F(1, 7) = 6.4$, $p < .05$, and for the D condition ($F = 12.7$, $p < .01$). The influence of interference of RTs to new signals cannot be predicted from its influence on RTs to repeated signals. The latter is in fact practically nonexistent, as can be seen from the means in Table 3. For the C condition, the small interference effects on both categories of RTs which are suggested by the means in Table 3 are nonsignificant.

The conclusion is that under the two conditions where an interference effect on RT could be obtained, it affected mainly RT to new signals.

The reason why no interference effect could be obtained under the C condition is probably that performance under that condition is permanently hampered by interference from the very familiar relationship which defines Cond. D so that no short-term interference could be effective. In this experiment, the largest interference effect was obtained under Cond. P, and this suggests the hypothesis that short-term interference can best be observed with systems where no very dominant S-R pairing exists.

EXPERIMENT II

The data to be presented next were collected during a larger experiment which is described elsewhere (Bertelson & Joffe, 1962). The main purpose of that experiment was to study the effect of two tranquilizers on prolonged performance in a serial four-choice task. After the main task, part of the *S*s were given four runs of 50 responses each, first with a straightforward S-R relationship, which had been used in the main experiment, then with a different presumably less compatible relationship. These conditions were introduced mainly to see if a less compatible task would be more sensitive to drug effects. The results from the four runs under the compatible condition were analyzed in terms of the repetition effect. They showed (*a*) that the effect is present in a four-choice task as well as in the two-choice one and (*b*) that it is not affected by the drugs. The measurements taken under the second, less compatible, condition now offer an opportunity to validate the results of Exp. I on a somewhat different task and with a larger number of *S*s.

Method

Apparatus.—The task, described more fully elsewhere (Bertelson & Joffe, 1962), consisted of pressing one of four keys in response to the appearance of the numerals 1–4 on a Nixie numerical indicator. The display was switched off by the response, and the next signal appeared either when the key was *released* or 150 msec. after it was *pressed*, according to whichever gave the longer interval. The sequence of signals was random and was controlled by a punched tape programmer. The RTs to new and to repeated signals were cumulated separately on two decatron chronoscopes, while magnetic counters recorded the number of signals in each category.

Subjects.—A total of 33 male students, aged 21–30, volunteered to take part in the experiment. For the two sessions, they were paid a fixed wage of 500 Belgian francs ($10.00) plus a bonus ranging from 0 to 200 francs, according to their speed and accuracy.

Procedure.—The data to be reported here were gathered at the end of two long sessions, during which the *S*s had worked uninterruptedly for 30 min. on the first task and 40 min. on the second one. This was done with a straightforward S-R correspondence, the first key on the left corresponding to the numeral 1, the second to the numeral 2, etc. (Cond. 1234). At the end of the second session, about 5 min. after having completed the 40 min. of uninterrupted work, the *S*s did four short runs of 50 responses with the same S-R relationship, and then four runs with a different relationship, the numerals corresponding to the keys being, from left to right: 3, 2, 4, and 1 (Cond. 3241). For the second session, the *S*s were divided into three groups, on the basis of speed and accuracy during the first one. One group received 1,200 mg. meprobamate, another 1,200 mg. emylcamate, and the third a placebo, 1 hr. before the experiment.

Results

The relevant data for each *S* are mean RT to new signals and mean

TABLE 4

ANALYSES OF VARIANCE AND COVARIANCE: EXP. II

Source	Variance Analyses				Covariance Analysis of y		
	df	*MS*			*df*	*MS*	*F*
		x^2	xy	y^2			
*S*s							
Between groups (G)	2	218	285	398	2	63	
Within groups	30	172	160	318	30	169	
Conditions (C)	1	1.532**	4.284	11.975**	1	1.564	74.5**
C × G	2	9	12	15	2	1	
Residual	30	28	31	55	29	21	
Total	65				64		

Note.—y = RT to new signals; x = RT to repeated signals.
** $p < .01$.

TABLE 5

RTs in Msec. for Exp. II

Cond.	Signals		Difference New − Repeated
	New	Repeated	
1234	565	422	143
3241	834	518	316
Difference (3241) − (1234)	269	96	173

RT to repeated signals in the four runs under Cond. 1234 and 3241.

Analysis of covariance was again applied to RTs to new signals, the RTs to repeated signals serving as predictor (Table 4). A highly significant ($p < .01$) conditions effect appears. Since there is no interaction between conditions and drug groups, all groups were pooled to give the average RTs given in Table 5. The RTs to new signals are much more affected by the modification of the S-R relationship than the RTs to repeated signals. The latter are nevertheless significantly affected, as the analysis of variance shows (Table 4). These results clearly confirm those of Exp. I.

Discussion

The results do not give unequivocal support to either of the alternative hypotheses considered in the introduction. The fact that one category of RTs is much more affected by S-R relationships than the other is inconsistent with the second hypothesis (same mechanism with facilitation in the case of repetitions). But the fact that RTs to repeated signals are affected, although to a much lesser degree, does not permit acceptance of the first hypothesis (different mechanism). The dichotomy was probably too simple. A more sophisticated interpretation, integrating features of the two former ones, can now be proposed. Reactions to new signals involve processes the duration of which depends on S-R relationships. Reactions to repeated signals can be organized via a shorter process, with duration independent of S-R relationships. But this economical mechanism does not work reliably, so that the other, less economical, mechanism is called forth in a definite proportion of cases.

It is now customary to conceive choice RT as resulting from some sort of progressive classification process (Falmagne, 1961; Hick, 1952; Welford, 1960). The uncertainty about the response to give is progressively reduced by asking dichotomous questions about the stimulus. The optimum strategy consists of asking questions such that probability of a positive and a negative answer is as near .5 as possible. Several deviations from this optimal strategy can occur. (*a*) One type of deviation consists of asking questions with one very probable answer. An extreme case of this type of strategy is serial inspection, where questions of the type "Is this stimulus a 3?" are asked until a positive answer is obtained, or until a negative answer has been given for all alternatives but one. (*b*) Another type of deviation consists of asking redundant questions, the answer to which can be inferred from those already asked. The strategy considered by Welford (1960) under the name "serial classification with check on accuracy" is an example: it consists of checking, whenever a negative answer has been obtained for one question, that the opposite question is answered positively. Welford has shown that if such a strategy is hypothesized, the effect of frequency unbalance on *binary* choice RT can be accounted for. This effect is difficult to explain in case of an optimal strategy. (*c*) Still another type of deviation consists of asking irrelevant questions, i.e., which are of no value in choosing a response.

The effect of S-R relationships on RT can be interpreted as reflecting changes in the classification strategy. For the same amount of information to be transmitted, poor S-R relationships would involve more classification steps than more compatible ones. This is exactly how Deininger and Fitts (1955,

p. 318) interpret compatibility, if "number of operations" employed "to effect transformation" is the same as "number of classification steps." The interference from the preceding performance under a different S-R relationship could be attributed to the persistence of the classification strategy, which was used under that condition.

A classification system which would show a repetition effect would be one which, whatever the number of alternatives, would begin by asking whether the stimulus is identical to the preceding one (this step can be called the "repeat question"). For a repeated stimulus, the decision can thus be taken in one step. For a new stimulus, in the case of choice between more than two alternatives, it is obvious that more than one step will be necessary whatever the classification strategy. In the case of two alternatives, the system will take more than one step for new signals only if some sort of Type *b* (see above) deviation from the optimal strategy is involved: after finding out that the stimulus is not the same as the preceding one, the system checks to see if it is the other one, or even proceeds to classify it, as though no information had yet been gained about it.

If poor compatibility implies asking too many questions, but if for repeated signals the decision is always reached after the "repeat question," the RT to new signals only will be affected.

The fact that the RT to repeated signals is also affected, although to a much lesser degree, would mean that the "repeat question" is not always asked first. The mechanism of this question necessarily involves some memory device where the trace of the preceding stimulus is stored. If the trace undergoes a decay, the "repeat question" cannot be asked reliably. This hypothesis at the same time explains that the repetition effect is reduced when the time interval since the last response is increased (Bertelson, 1961).

REFERENCES

BERTELSON, P. Sequential redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, 1961, **12**, 90–102.

BERTELSON, P., & DAVIDSON, A. A simple two-channel programming unit permitting the control of conditional probabilities. *Quart. J. exp. Psychol.*, 1959, **9**, 180–184.

BERTELSON, P., & JOFFE, R. L'influence du méprobamate et de l'émylcamate sur la performance humaine dans une tâche sériale prolongée. *Psychopharmacologia*, 1962, **3**, 242–253.

DEININGER, R. L., & FITTS, P. M. Stimulus-response compatibility, information theory and perceptual-motor performance. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe: Free Press, 1955. Pp. 316–341.

FALMAGNE, J. C. Codage de l'information et temps de réaction de choix: II. Les canaux bruyants. *Cah. Cent. Rech. Operat.*, 1961, **3**, 133–157.

FITTS, P. M., & DEININGER, R. L. S-R compatibility: Correspondence among paired elements within stimulus and response codes. *J. exp. Psychol.*, 1954, **48**, 483–492.

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

WELFORD, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–230.

(Received July 5, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 485–489

FAILURES TO OBTAIN MEDIATED GENERALIZATION EFFECTS IN EYELID CONDITIONING[1]

G. ROBERT GRICE, HOWARD J. SIMMONS, AND JOHN J. HUNTER

University of Illinois

4 experiments employing differential conditioning of the eyelid response were conducted to investigate the phenomenon of mediated generalization. The intended mediating responses were instructed manual or verbal reactions. 3 of the experiments involved implicit representation of these responses initiated by the onset of the CS. In the 4th experiment the mediating response occurred just prior to the CS in order that the response-produced cues might form a favorable conditioning interval with the UCS. None of the experiments provided evidence for the operation of mediational effects. Experiments of this kind have yielded positive evidence only when the mediating response was overt and in close temporal proximity to the mediated response elicited.

The present paper represents an extension of research reported by Grice and Davis (1958, 1960) which indicated the operation of a mediated generalization effect in human eyelid conditioning. The experimental paradigm of this research is a differential conditioning procedure involving three stimuli, one of which is the CS. The *S*s then make two instructed "mediating" responses. One of these is made to the CS and to one of the negative stimuli; the other to the second negative stimulus. Evidence for mediation consists of asymmetrical generalization, in which there are more conditioned eyelid responses to the negative stimulus to which the mediating response is the same as to the CS than to the one for which it is different. In all of the experiments which have yielded positive results so far, the mediators were overt, instructed, manual reactions made as reaction time responses to the three stimuli. However, in most of the theoretical accounts of mediated generalization, the assumed mediators are implicit. The first two experiments reported here were designed to ascertain whether implicit states, based on the same mediators previously used, would produce the generalization effect. The third experiment, while utilizing an overt verbal response, turned out in such a way that it is best interpreted as related to the first two. The fourth experiment involved the overt manual responses of the previous research, but an attempt was made to manipulate the temporal relation between these responses and the UCS. It was an investigation of the hypothesis that mediational effects could be increased by bringing the mediating response and the UCS into a relation approximating a favorable CS-UCS interval.

EXPERIMENTS I, II, AND III

Grice and Davis (1958) reported one experiment in which the manual responses were not made until after the conditioning trial was over, thus testing the mediating efficacy of a set to respond. While the experiment did not produce a significant result, there was some asymmetry in the predicted direction. Experiment I is a similar experiment, but with an effort to

---

[1] The present research was supported by Grants G-4416 and G-14223 from the National Science Foundation.

intensify the effects of set. This was done by instructing *S* to perform a reaction time response to a click which was coincident with the termination of the CS. Since only anticipatory CRs were scored, all conditioned and generalized responses occurred prior to the overt response.

Experiment II is another approach to the study of implicit representation of the mediating responses. The experiment began, as in the previous studies, with the *S*s making manual RT responses to the three stimuli. However, during a block of 60 trials in which generalization effects were assessed, they were instructed to discontinue these responses, but on each trial to "think about" and implicitly rehearse the correct response.

Experiment III was an attempt to utilize verbal responses rather than the manual lever pushing responses of the previous experiments. The *S*s said one word as an RT response to the CS and to one negative stimulus, and another word to the second negative stimulus. However, the reaction latencies of the verbal responses turned out to be longer than for the manual responses, and most of the responses occurred after the end of the CS-UCS interval. Thus, in effect, the experiment was actually very similar to Exp. I.

Method

Subjects

The *S*s were undergraduate men and women who served as part of a requirement for a course in introductory psychology. There were 12 men and 12 women in Exp. I, 32 women in Exp. II, and 18 women in Exp. III. All *S*s included made at least 10 CRs in Trials 41–100.

Apparatus

The apparatus and general procedures have been previously described (Grice & Davis, 1960). Briefly, the apparatus consisted of conventional eye-blink recording equipment and programing equipment for the presentation of three stimuli and an air-puff UCS in a 100-trial sequence. The negative stimuli were always tones of 240 and 1,900 cps. The CS was either a light (Exp. I) or a tone of 850 cps (Exp. II and III). In Exp. I and III, tones were presented at a 70-db. loudness level with respect to an average threshold. In Exp. II, which was actually conducted last, it was necessary to change the earphones. The new phones were Telephonics, Type TDH-39, calibrated by Grayson-Stadler. The tones were nominally at an 80-db. loudness level. They were judged equal in loudness to a 1,000-cps tone produced by a voltage for 80 db. SPL according to the calibration. However, since calibration was with a 6-cc chamber and the ear pads employed in the experiment had a volume of 25 cc, the actual SPL would be somewhat attenuated. The UCS was a puff of air of 2 psi, and the CS-UCS interval was .5 sec. Responses resulting in pen deflections of 1 mm. or more in the interval 150 to 500 msec. were scored as CRs in Exp. I and III. In Exp. II, provision was made for scoring responses to the negative stimuli in the interval 200 to 800 msec.

Procedure

Experiment I.—The right and left hand lever responses were the same as previously described. Each *S* made one of these responses to the CS (light) and to one of the negative tones, and the other response to the other tone. Complete counterbalancing of right and left responses and tones was employed as before. Coincident with the termination of the light and the tones was a loud click produced by a condenser discharge. The *S* was instructed to operate the proper lever as quickly as possible when he heard the click. The UCS was paired with the light only. There were 100 trials, the three stimuli being presented in an irregular order. There were 40 conditioning trials with the CS, and 30 presentations of each negative stimulus. The last 60 trials consisted of 20 presentations of each stimulus.

Experiment II.—The experiment began in the same way as Exp. I of Grice and Davis (1960). The middle tone was the CS, and the two manual reactions were made as reaction time responses to the three tones. After Trial 40, *S* was told to stop snapping the switches, but for each tone to think to herself what the correct response was. She was told that she would be tested later.

Experiment III.—Verbal responses were employed here rather than the manual reactions. Half of the *S*s were told to say "High" as soon as they heard either of the two higher tones, and "Low" when they heard the low tone. For counterbalancing purposes, the other half was instructed to say "Low" to the two lower tones and "High" to the high tone. Actually the middle tone (the CS) was the same for both groups. The tones for which the verbal response was the same as to the CS were combined in analysis for the purpose of evaluating any mediational effect, and those to which the verbal response was different were similarly combined. Reaction times of the verbal responses were recorded by means of a voice-key.

Results

For all three experiments, data are presented in terms of number of CRs during Trials 41–100 which included 20 presentations of each stimulus. Evidence for the mediation effect would be more responses to the "Same Response" tone than to the "Different Response" tone. Mean number of CRs to each stimulus for the three experiments are presented in Table 1. No marked tendency for the predicted asymmetry is evident in any of the experiments. As in the previous experiments, analyses of variance were performed on the responses to the negative stimuli, treating the counterbalanced variables as between-*S*s effects and the generalization effect as a within-*S*s effect. The generalization effect was significant in none of the experiments—all three of the *F* values being less than unity.

TABLE 1

Mean Number of CRs to Each Stimulus in Exp. I, II, and III

Stimulus	Exp. I		Exp. II		Exp. III	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Same response	3.63	4.41	9.75	5.56	10.06	5.15
CS	14.13	3.69	8.34	5.15	13.50	5.26
Diff. response	3.50	4.12	9.81	5.59	9.94	6.16

In Table 1, the means for the negative stimuli in Exp. II are not comparable with the mean for the CS, since they include responses in the interval 200 to 800 msec. rather than to the onset of the UCS.

In Exp. III, the latencies of the verbal responses were recorded. The mean values for the "Same Response," conditioned, and "Different Response" stimuli were .87, .89, and .89 sec., respectively. Since the CS-UCS interval was .5 sec., this means that a high proportion of the verbal responses occurred after the end of the trial. This is quite different from the results obtained by Grice and Davis with manual responses.

In evaluation of these negative findings, it is well to keep in mind the precision of the within-*S* comparisons which is not indicated by the *SD*s of Table 1. The .95 confidence limits for difference in the predicted direction between the mean numbers of responses to the two negative stimuli are as follows: Exp. I, 1.15; Exp. II, 1.31; and Exp. III, 1.92.

Experiment IV

According to conventional mediation theory, the cues produced by a response are supposed to serve as stimuli for further responses and thus to provide a basis for such phenomena as mediated generalization. In the Grice and Davis experiments yielding positive evidence, the mediating responses and the, presumably, mediated eyelid response both occurred in close temporal proximity within a 500-msec. CS-UCS interval. In the three negative experiments reported above, either there was no overt mediator at all, or it did not occur until the end of the trial. In view of our knowledge concerning the CS-UCS interval in conditioning, one might expect that neither of these procedures would be particularly effective. Presumably, the mediating response should precede the UCS by some favorable conditioning interval. Experiment IV represents one approach to approximating such an arrangement.

Method

The three tones employed as stimuli and the UCS were the same as in Exp. III above. However, in this experiment the tones were presented twice on each trial. There were two 500-msec. presentations separated by an interval of 1 sec. The *S* was instructed to snap the correct switch as quickly as possible to the first presentation of the tone. The UCS occurred at the termination of the second presentation in the case of the CS. Eyelid responses were recorded during the interval 150–500 msec. of the second presentation. The experimental design involving the right and left hand responses and the three tones was the same as in the previous experiments. The *S*s were 36 women from the same population as above.

Results

No significant difference was obtained in the number of CRs to the two negative stimuli. The mean numbers of blinks to the three stimuli on Trials 41–100 were as follows: Same Response Stimulus, 6.19, $SD = 3.72$; Different Response Stimulus, 6.14, $SD = 3.71$; CS, 10.11, $SD = 4.31$. In the analysis of variance of responses to the negative tones, the value of F for the generalization effect was less than unity. The .95 confidence limit for a difference in the predicted direction is 1.37 responses.

There is no assurance that the 1-sec. interval between the two stimulus presentations is the best for such an experiment. The reasoning behind its choice was that almost all of the lever responses would be complete before the second stimulus, and still it would be short enough that traces from the response would be present during the CS-UCS interval of the second stimulus. Of course, the actual interval between the response and the UCS was variable, depending on the reaction time of *S*. While this is clearly not the best possible arrangement from the point of view of conditioning, it does permit some evaluation of a possible relation between the temporal location of the manual response and the probability of a CR. The overall mean RT was 703 msec. The mean on trials on which a CR occurred was 713 msec., and it was 690 msec. on trials with no CR. This difference was not significant, $t = 1.04$, $p > .3$. One possibility would be that *S*s responding slower on the average might be more inclined to show the mediational effect since longer RTs would result in a shorter interval between the response and the UCS. The data do not support such a conclusion, however. There were 16 *S*s who made more CRs to the Same Response Stimulus than to the Different Response Stimulus, and 17 who made more to the Different Response Stimulus. The mean RTs of these two groups were almost identical—711 and 709 msec., respectively.

DISCUSSION

The evidence from the present four experiments is overwhelmingly negative. The results of Exp. I together with the previous finding of Grice and Davis (1958) provide rather convincing evidence that a set to respond differentially to two stimuli which is initiated at the beginning of the CS-UCS interval is not a sufficient condition for the mediated generalization of the conditioned eyelid response. The results of Exp. III do not warrant the conclusion that verbal responses cannot serve as mediators as the manual responses apparently did in previous experiments. It is still possible that a more favorable temporal arrangement may produce positive results with such responses. The long reaction times obtained for these responses, resulting in most of the responses occurring after the trial, suggest that the results support the conclusion of Exp. I.

Experiment II demonstrated that another type of implicit representation of differential response, which was also initiated at the beginning of the CS-UCS interval, was not a sufficient condition for generalization effects. Of course, one obvious difficulty in the interpretation of this experiment is that we have no polygraphic measures of thought. We have no information as to exactly when it occurred or concerning its consistency from trial to trial.

The negative finding of Exp. IV appears to be definitely embarrassing to conventional mediation theory. There can be little doubt that traces of the overt, differential responses existed at the onset of the CS and during the CS-UCS interval. One seems forced to conclude that a very reasonable prediction of mediation theory was not confirmed.

A fairly accurate summary of this research program to date, is that mediational effects have been obtained only when the mediating response was overt, and when it occurred in close temporal proximity to the mediated response. This is by no means a prediction readily made from existing theory concerning mediating processes. It also represents the condition representing the greatest danger from facilitative artifacts of the type investigated by Grice and Davis (1960). It is the writers' opinion that the next step in these investigations should involve further manipulation and detailed examination of the temporal relations among these variables.

REFERENCES

GRICE, G. R., & DAVIS, J. D. Mediated stimulus equivalence and distinctiveness in human conditioning. *J. exp. Psychol.*, 1958, 55, 565–571.

GRICE, G. R., & DAVIS, J. D. Effect of concurrent responses on the evocation and generalization of the conditioned eyeblink. *J. exp. Psychol.*, 1960, 59, 391–395.

(Received July 7, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 490–494

PERCEPTION OF THE POSTURAL VERTICAL IN NORMALS AND SUBJECTS WITH LABYRINTHINE DEFECTS[1]

BRANT CLARK

San Jose State College

AND ASHTON GRAYBIEL

United States Naval School of Aviation Medicine, Pensacola, Florida

9 normal men and 10 men with vestibular defects were studied in a lateral tilt chair for 30 successive settings to the postural vertical. A significant decrease in the average error was found for both groups. The normal Ss exhibited smaller average errors; however, the differences were small particularly after 15 trials and were not statistically significant.

The ability of normal *S*s to set themselves to the postural vertical in the absence of a visual framework is remarkably accurate, the threshold values being of the order of 1–3° (Clegg & Dunfield, 1954; Fleishman, 1953; Garten, 1920; Mann, Berthelot-Berry, & Dauterive, 1949; Pearson & Hauty, 1960; Solley, 1960; Witkin & Asch, 1948). There has been much speculation in the literature with respect to the specific perceptual factors which are involved, and practically all possible sensory mechanisms have been listed as contributing to the process, notably tactual, kinesthetic, and vestibular cues (Garten, 1920; Mann et al., 1949; Pearson & Hauty, 1960). Relatively little clear evidence has been obtained regarding the specific nature of the sensory mechanisms, however. Tactual cues have been shown to be important by several investigators, notably Mann et al. (1949) and Solley (1956). It has been generally assumed that the vestibular mechanism plays a dominant role in setting to the postural vertical in the absence of visual cues, and there is some evidence that *S*s without vestibular cues do not perform this task as well as normals. Garten (1920) reported that deaf *S*s were less accurate than normals, but the differences were small. Padden (1959) studied the ability of blindfolded, deaf (and presumably avestibular) *S*s to orient themselves to the vertical by swimming to the surface after being dumped into the water from a special platform. Some of the deaf *S*s became completely disoriented and even swam down instead of up, but in general their performance was remarkably like that of normals. In this connection it is of interest to note that Brown (1961) found gross errors in blindfolded, normal *S*s who were rotated under water and then asked to point to the surface.

Although there may be some question regarding the specific sensory processes involved in setting to the postural vertical, there is considerable evidence to indicate that accuracy increases with practice. Solley (1956, 1960) presents practice curves for men which show marked improvement for about 15 trials with slight improvement thereafter up to 30 settings of the chair to vertical. Evidence of learning has also been reported by Garten (1920), Fleishman (1953), and Pearson and Hauty (1959, 1960). It would appear that basic information regarding the sensory processes involved in the improvement with practice in setting to the postural vertical would be obtained by comparing the performance of normal and nonnormal *S*s in this situation. Therefore, this study was designed to compare normal *S*s and *S*s with defective vestibular function with respect to the effects of practice on their ability to set themselves to the postural vertical.

---

[1] This research was conducted under the sponsorship of the Office of Life Science Programs, National Aeronautics and Space Administration (Grant No. R-37). Opinions or conclusions contained in this report are those of the authors and do not necessarily reflect the views or endorsement of the Navy Department.

We wish to express our appreciation to the *S*s who participated, especially Ensigns Nugent and Thurman who assisted in the collection of the data, and to the members of the Statistical Department for their assistance in analyzing the data.

METHOD

Apparatus.—The apparatus consisted of a large tilting chair which could be controlled by *E* or *S* independently. The chair had a wooden seat for these observations and was tilted, with the center of rotation at seat level, by a hydraulic motor controlled by either *E* or *S* turning a knob. The angular velocity was variable depending on the pressure applied to the knob, the typical recovery velocity varying between 2 and 3 rpm. This was well above the optimum level reported by Clegg and Dunfield (1954). The *S* was strapped firmly in position in the chair with seat belt and shoulder straps, and his head was held firmly in position by a rigid head rest. The position of the chair could be read on a dial to an accuracy of ½° by *E* who sat at a table directly behind *S*. The room in which the chair was located was darkened during all of the tests, and *S*'s eyes were covered with eye patches.

Subjects.—Nine control *S*s and 10 *S*s with vestibular defects took part in the experiment. Medical examination revealed that all were in good general health. All of the controls had normal hearing, exhibited normal responses to passive rotation, and perceived the oculogravic illusion. All of the nonnormal *S*s acquired their deafness in childhood as a sequela of meningitis. Seven of the 10 had some residual hearing when exposed to high intensity noise. Although none exhibited definite nystagmus as a result of caloric stimulation (ice water), questionable nystagmic beats (not sustained) were recorded in 3. Four *S*s including these 3 also perceived the oculogravic illusion, indicating at least some functioning of the otolith apparatus. The remaining 6 did not perceive the illusion.

Procedure.—The experimental procedure was similar to that used by Solley (1956) but differed in several important respects. Each *S* was instructed regarding his task and was given several practice trials in controlling the chair while he stood to one side of it and operated the controls on signal from *E*. But no preliminary practice was given in the chair. He was told that his task was to set himself to the upright with respect to the outside world on signal from *E*. The *S*'s eyes were then bandaged, and the room was darkened. The series of trials was given as follows: The chair was in a vertical position before *S* was tilted and *E* signaled, "upright." The *S* sat in the upright position for 5 sec., and then the chair was tilted 30° to the right. On reaching the tilted position, *E* signaled to *S* who immediately aligned himself to the vertical. The *S* signaled when he felt he was in an upright position, and *E* recorded the magnitude and direction of the error. The *S* was then returned directly to the physical upright for 5 sec., after which he was tilted 30° to the left. Tilts to the right and left were alternated, and 30 trials were given each *S* with a 60-sec. rest period between each series of 10 trials. After the test series, *E* asked *S* to report on his subjective experiences and in particular to indicate which cues were judged to be helpful in accomplishing his task.

RESULTS

Plots of the median error for each of the 30 practice trials for the normal *S*s and *S*s with vestibular defects are presented in Fig. 1. These data show that with practice there is a reduction in the median error in setting the chair to the postural vertical. Comparable data from a study by Solley (1956) show a striking similarity in the practice curves for the normal *S*s despite the fact that the method in this experiment was somewhat different. Solley's *S*s showed greater error

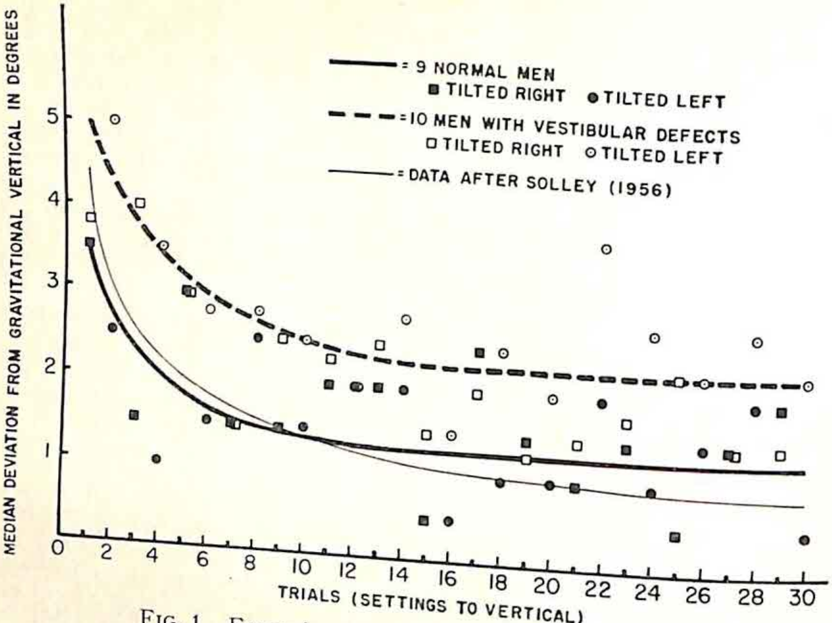


FIG. 1. Errors in setting to the postural vertical.

at the outset, continued to improve longer, and reached a somewhat lower error level. However, it is unlikely that these differences are significant. To test the significance of the differences among the successive blocks of trials, a trend analysis was performed (Edwards, 1950) after randomly eliminating 1 *S* from the nonnormal group to equalize the groups. The F (5, 5) for trials was found to be 11.00 ($p < .01$), indicating that the successive blocks of trials differ significantly. The different Groups of *S*s × Trials interaction gave an $F\ (5, 80) = 1.7$, $p > .05$, showing no significant difference between the forms of the two curves.

The normal *S*s were superior to the nonnormal *S*s throughout the six blocks of trials (Table 1), but the differences were small particularly during the final trials and did not reach statistical significance in the analysis of variance, $F\ (1, 16) = 3.36$, $p > .05$. It should be noted that in the final three blocks of trials the differences between the two groups were less than a degree. These results confirm earlier findings by Garten (1920) and are similar to Solley's (1960) in showing that *S*s can improve with practice in setting themselves to the postural vertical in spite of sensory limitations which result in initial large errors.

One additional point should be made regarding the errors of *S*s with defective vestibular function. The

TABLE 1

MEAN ERROR IN DEGREES IN PERCEPTION OF POSTURAL VERTICAL IN BLOCKS OF FIVE TRIALS

Group	Trial Blocks					
	1–5	6–10	11–15	16–20	21–25	26–30
Nonnormal ($N = 10$)	4.4	2.8	2.8	2.2	2.6	2.4
Normal ($N = 9$)	2.6	1.9	1.7	1.6	1.7	1.7
Difference	1.8	0.9	1.1	0.6	0.9	0.7
Nonnormal ($N = 6$)	5.1	3.5	3.6	2.6	2.7	2.7
After Solley (1956)						
Left tilt	3.00	1.85	1.51	1.19	1.34	1.03
Right tilt	2.49	1.87	1.60	1.21	1.32	1.42

data reported in this study were collected on the first day of a 2-wk. testing schedule. Subsequent tests in the Pensacola Slow Rotation Room (Graybiel, Clark, & Zarriello, 1960) showed that 4 of the *S*s observed the oculogravic illusion, indicating some otolith function. The mean error by blocks of five trials for the remaining 6 *S*s shows that their errors were greater in all six blocks of trials and substantially greater than for the total group for about the first 20 trials (Table 1). These differences lend additional support to the notion that *S*s without otolith function can be expected to make marked improvement with practice.

An analysis of the subjective reports regarding the cues judged to be of importance in setting to the postural vertical indicated that they were much the same for the two groups. Both groups reported using pressure cues to the head, shoulders, thighs, and buttocks in making the settings. In general, these reports correspond to those reported by Solley (1956); nevertheless, there were some definite differences in subjective reports. Five nonnormal *S*s reported being seriously confused or lost with respect to their position whereas only 3 normal *S*s showed any feeling of confusion and these were very slight. These confusions may be considered to be comparable to those reported by Padden (1959) for deaf *S*s in water. The data also give support to the findings of others that settings tend to be made in the direction of tilt (Fleishman, 1953). This was true for both groups. For tilts to the *right*, the 9 normal *S*s made the following settings: right, 97; left, 32; and zero, 6; the 10 nonnormal *S*s set 83 right; 58 left; and 9 at zero. For the tilts to the left, the 9 normals set 84 left; 41 right; and 10 at zero while the nonnormal *S*s set 130 to the left; 16 to the right; and 4 at zero.

DISCUSSION

The two primary findings of this study were:

1. Both normal and nonnormal *S*s showed improvement with repeated trials in setting themselves to the postural vertical. This not only supports the results of similar studies of the postural vertical cited above, but it also supports the general principle that, in a wide variety of situations, practice in a perceptual task results in the perception becoming more veridical (Gibson, 1953).

2. The normal *S*s were superior to the *S*s with vestibular defects, but the differences were relatively small and not statistically significant. It should be pointed out that the method used was selected to facilitate high level performance. Although there is some difference of opinion in the literature regarding the factors which facilitate performance in this task (Pearson & Hauty, 1960), the following should be pointed out as factors which possibly contributed to "optimum performance" in this experiment: (*a*) The *S* passed through the vertical following each trial and was permitted to adjust the chair through the vertical in making his adjustment (Pearson & Hauty, 1960). (*b*) The *S* had knowledge of results for each setting since he was placed at the gravitational vertical following each trial. (*c*) A hard seat was used (Mann et al., 1949). (*d*) The head was held in a fixed position, and the body was firmly supported by the chair and straps (Fleishman, 1953; Solley, 1960; Werner, Wapner, & Chandler, 1951). (*e*) The *S* returned himself to the vertical immediately (Passey & Guedry, 1949; Pearson & Hauty, 1960).

Under these "optimum conditions" the normal *S*s exhibited smaller average errors at the outset than the group with vestibular defects, but after 15 settings the *S*s with defective vestibular function reduced their errors to approach very closely the performance of the normals.

These results show that normal vestibular function is not a prerequisite for an

S to set himself to the postural vertical with considerable accuracy, in the absence of a visual framework, particularly after practice. Subjective reports indicate that most of the *S*s were clearly aware of pressure cues from various parts of the body and that they used these cues in setting the chair. These cues appear to be adequate to insure good performance in the group with defective vestibular function. For example, 1 nonnormal *S* made a median error on the final 10 settings of only .5° which was the best performance in both groups while 1 normal *S* made a median error of 2.8° on the final 10 trials. At the same time it is of interest to note that the 4 nonnormal *S*s, who were judged to have some residual otolith function on the basis of having observed the oculogravic illusion, showed less average error during the final 10 trials (mean error—2.1°) than the 6 who did not perceive this illusion (mean error—2.7°). It is noteworthy that the individual with the smallest median error on the final 10 trials was judged to have no residual otolith function. Thus, the nonnormal *S*s learned to use perceptual data available to them for relatively good performance on this task in spite of the fact that they experienced confusions not found in normal *S*s. Furthermore, these results offer convincing evidence that the tilt chair, as used here, is not a valid indicator of vestibular function.

These results also suggest the strong possibility that pressure cues are of primary importance for normal *S*s in setting the chair to the gravitational vertical. The overall better performance of the normal *S*s particularly when compared with the small group of 6 nonnormals (Table 1) suggest that the otoliths may also play an important role especially in finer adjustments to the vertical without practice.

REFERENCES

BROWN, J. L. Orientation to the vertical during water immersion. *Aerospace Med.*, 1961, **32**, 209–217.

CLEGG, W. C., & DUNFIELD, N. M. Nonvisual perception of the vertical: II. Lateral plane. *Canad. J. Psychol.*, 1954, **8**, 80–86.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1950.

FLEISHMAN, E. A. Perception of body position in the absence of visual cues. *J. exp. Psychol.*, 1953, **46**, 261–270.

GARTEN, S. Über die Grundlagen unserer Orientierung im Raum. *Abh. Sachs. Ges. (Akad.) Wiss.*, 1920, **26**, 433–510.

GIBSON, E. J. Improvement in perceptual judgments as a function of controlled practice or training. *Psychol. Bull.*, 1953, **50**, 401–432.

GRAYBIEL, A., CLARK, B., & ZARRIELLO, J. J. Observations on human subjects living in a "slow rotation room" for periods of two days. *AMA Arch. Neurol.*, 1960, **3**, 55–73.

MANN, C. W., BERTHELOT-BERRY, N. H., & DAUTERIVE, H. J., JR. The perception of the vertical: I. Visual and non-labyrinthine cues. *J. exp. Psychol.*, 1949, **39**, 538–547.

PADDEN, D. A. Ability of deaf swimmers to orient themselves when submerged in water. *Res. Quart. Amer. Ass. Hlth. Phys. Educ. Recr.*, 1959, **30**, 214–226.

PASSEY, G. E., & GUEDRY, F. E., JR. The perception of the vertical: II. Adaptation effects in four planes. *J. exp. Psychol.*, 1949, **39**, 700–707.

PEARSON, R. G., & HAUTY, G. T. Adaptive processes determining proprioceptive perception of verticality. *J. exp. Psychol.*, 1959, **57**, 367–371.

PEARSON, R. G., & HAUTY, G. T. Role of postural experiences in proprioceptive perception of verticality. *J. exp. Psychol.*, 1960, **59**, 425–428.

SOLLEY, C. M. Reduction of error with practice in perception of the postural vertical. *J. exp. Psychol.*, 1956, **52**, 329–333.

SOLLEY, C. M. Influence of head tilt, body tilt, and practice on reduction of error in perception of the postural vertical. *J. gen. Psychol.*, 1960, **62**, 69–74.

WERNER, H., WAPNER, S., & CHANDLER, K. A. Experiments on sensory-tonic field theory of perception: II. Effect of supported and unsupported tilt of the body on the visual perception of verticality. *J. exp. Psychol.*, 1951, **42**, 346–350.

WITKIN, H. A., & ASCH, S. E. Studies in space orientation. III. Perception of the upright in the absence of a visual field. *J. exp. Psychol.*, 1948, **38**, 603–614.

(Received July 11, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 495–500

EXTINCTION IN A RUNWAY AS A FUNCTION OF ACQUISITION LEVEL AND REINFORCEMENT PERCENTAGE[1]

WINFRED F. HILL AND NORMAN E. SPEAR

Northwestern University

Level of acquisition training (8, 16, 32, 64, and 128 trials) and percentage of food reinforcement (100% and 50% random) were combined factorially for 100 rats in a straight alley. Acquisition was by massed training and was followed by 28 trials of massed extinction on one day and 12 on the next. Running speed in late acquisition was an increasing function of training level, with no significant effect of percentage. Running speed early in extinction was also an increasing function of training. This effect disappeared as extinction progressed for the 100% but not for the 50% groups. These results do not support the hypothesis, suggested by previous studies, that resistance to extinction is an inverted U function of degree of training.

Hullian theory predicts that resistance to extinction will increase as the number of reinforced acquisition trials increases (Hull, 1952). The bulk of the support for this prediction has come from Skinner box studies (e.g., Harris & Nygaard, 1961; Perin, 1942). However, some runway studies support the same conclusion. Lawrence and Miller (1947), after explaining away some earlier contrary findings, demonstrated greater resistance to extinction of a running response after 16 reinforced trials than after 8. Similarly, Lewis and Cotton (1959) found greater resistance to extinction after 60 trials on random 50% reinforcement than after 16 trials.

Other recent data, however, have suggested that resistance to extinction may not be a monotonic increasing function of number of acquisition trials. North and Stimmel (1960) found faster running during extinction after 45 continuously reinforced trials in a runway than after either 90 or 135 trials. Taken in conjunction with the other runway studies mentioned above, this suggests that as acquisition training is continued beyond some point, resistance to extinction may reach a peak and then decline. The possibility of some function other than a monotonic increasing one is also suggested by a variety of other studies, human and animal (e.g., Capaldi, 1958; Lewis & Duncan, 1958; Murillo & Capaldi, 1961).

It is noteworthy that the runway studies have used two or at most three acquisition levels, so that the hypothesis of an inverted U function has not been adequately tested within a single experiment. The present study attempts to fill this gap by comparing extinction after five different acquisition levels with both continuous and partial reinforcement.

METHOD

Subjects.—The *S*s were 100 experimentally naive, female albino rats of the Sprague-Dawley strain, 74–77 days old at the beginning of experimental training. Eight additional *S*s were discarded; 3 because they would not eat the reward pellet within 2 min. for three consecutive trials, 3 because 2 min. elapsed from the starting to the stopping of either clock on three consecutive trials, and 2 because of *E* error.

---

[1] This research was supported by Grant G-8706 from the National Science Foundation.

Apparatus.—The *S*s were trained in an enclosed straight runway. This consisted of a start box 12 in. long and 5 in. wide, a stem 30 in. long and 3½ in. wide, and a goal box 12 in. long and 5 in. wide. The start box and stem were 4 in. high, with floors of wood and sides of clear plastic covering wood. There was a ¾-in. step-down from the stem to the goal box making the goal box 4¾ in. high. The goal box had wooden sides and floor painted flat black. The inside walls of the start box and stem were painted white and the floors were gray. Sliding doors separated the start box from the stem and the stem from the goal box. A .01-sec. timer was started automatically when the start-box door was opened and stopped when *S* passed through a photocell beam 6 in. from the start-box door. The breaking of this first beam activated a second .01-sec. timer which was stopped when *S* passed through a second beam located 12 in. from the first. Thus, the first timer measured starting time while the second measured running time. The second beam was 12 in. from the goal box.

Prehandling.—Each *S* received six daily 3-min. sessions of prehandling, the last one 48 hr. prior to the beginning of experimental training. During each session *S* was allowed to explore a large unpainted wooden box, presented with four of the pellets later to serve as reward, and picked up and replaced at least five times by *E*. Pellets uneaten were returned with *S* to the home cage along with a fifth pellet. A once-daily feeding schedule began on the first day of prehandling and was maintained throughout experimental training. The ration was 10 gm. of finely ground Purina lab chow and was presented 60 min. after the start of prehandling or experimental training. To maintain a constant ration for all *S*s, the amount of ground chow presented was adjusted according to the number of reward pellets consumed (if any) on a given day.

Experimental training.—On each trial, after placement into the start box, the door was opened as soon as *S* was oriented toward it. After *S* entered the goal box, the door was closed behind it and *S* was permitted to remain there for 10 sec. From the time of removal from the goal box to placement in the start box for the next trial, 30 sec. elapsed during which *S* was placed in its carrying cage (also its home cage) where water was always available. If on any acquisition trial *S* failed to stop a given timer within 2 min., *S* was removed from the runway and placed in the goal box for the usual 10-sec. period followed by the 30-sec. intertrial interval. If *S* failed to stop a given timer within 2 min. during extinction, it was manually removed from the runway and placed in its carrying (home) cage for the usual 30 sec. If, in acquisition or extinction, *S* failed to stop the second clock within the 2-min. limit, its running time for that trial was recorded as 120 sec. If *S* failed to stop the first clock within the limit, both start and running times were recorded as 120 sec. On rewarded acquisition trials, one .045-gm. Noyes pellet was available in a glass container in the goal box. On nonrewarded trials (extinction or acquisition), the pellet container was not present in the goal box.

Design.—Number of training trials in acquisition (8, 16, 32, 64, or 128) was orthogonally varied with percentage of reinforced acquisition trials (50 or 100) using 10 independent groups of 10 *S*s each. Acquisition was completed in one session for groups receiving 8, 16, or 32 trials, and 32 trials per day were given the other groups on successive days until completion of training. All groups were treated alike during extinction and were given 28 trials on the day immediately following the last acquisition session and 12 on the next day. Extinction differed from acquisition only with respect to the treatment of nonruns noted above and the absence of reward and container from the goal box.

RESULTS

Acquisition.—The acquisition curves for the two 128-trial groups are shown in Fig. 1. These curves show a suggestion of the tendency, previously noted by Goodrich (1959) and others, for the 100% group to run faster early in acquisition and the 50% group to run faster later in acquisition. However, the difference between these two groups on the last 4 trials of acquisition falls just short of significance, $t\,(18) = 1.95$, and that on the entire 32 trials of the last day of acquisition is even less significant (t of 1.25). In addition, the 100% curves shows a combination of decrement within days and reminiscence from one day to the next, a phenomenon also found in a different alley by Hill and Spear (1962).

Mean running speeds for the 10 groups on the last 4 acquisition trials are shown as the first point in Fig. 2 (100% groups) and Fig. 3 (50% groups). The increase in speed with number of acquisition trials, which

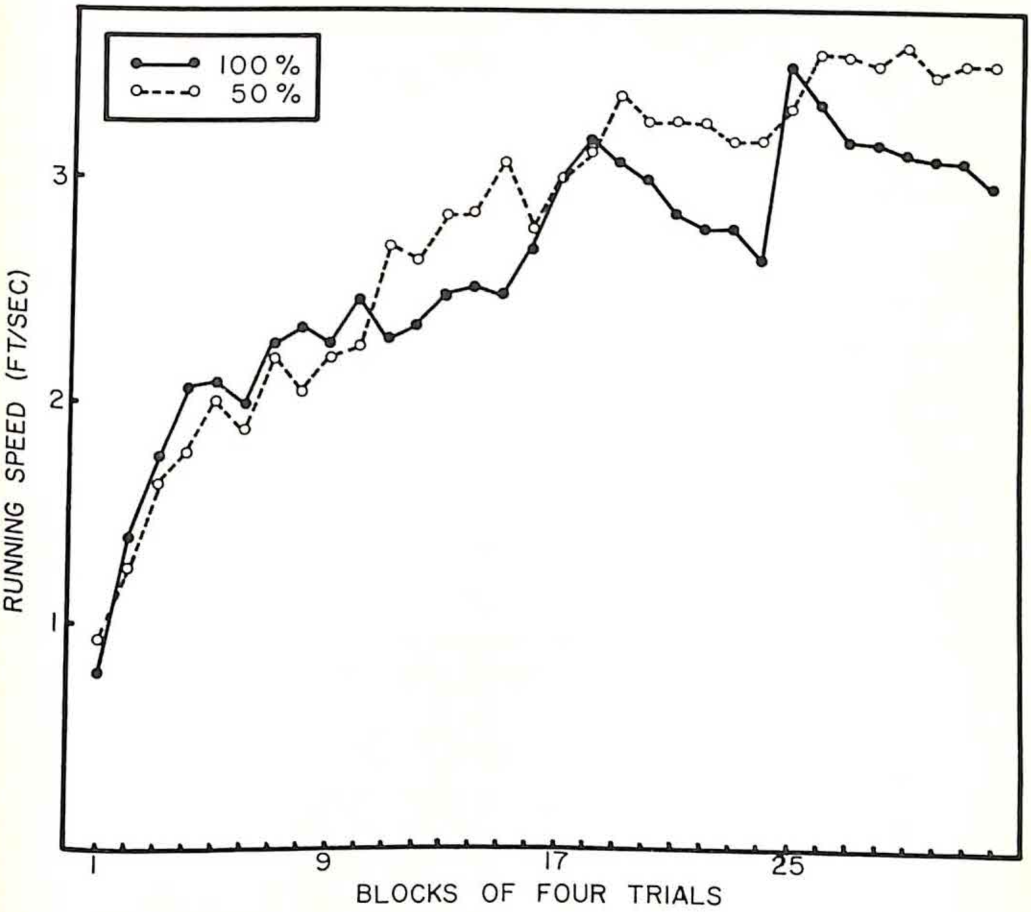


FIG. 1. Acquisition curves for the two 128-trial groups.

appears to be a log linear function, is highly significant, F (4, 90) = 24.62, $p < .001$. A test for multiple comparisons (Duncan's new range test; from Edwards, 1960) revealed apparently asymptotic performance by the 100% groups within 32 trials. That is, Groups 32-100, 64-100, and 128-100 did not differ among themselves but were significantly faster ($p < .01$) than 16-100 and 8-100. Within the 50% groups, *S*s given 64 or 128 trials were also equivalent in speed but differed from Groups 32-50, 16-50, and 8-50 ($p < .01$). The effects of percentage reward and of the interaction between number of trials and percentage are both nonsignificant, F (1, 90) = 1.36 and F (4, 90) = 1.86.

The starting latencies showed similar patterns to the running speeds, and detailed analyses of this measure are not reported. Two points, however, deserve mention. (*a*) The within-days decrement and the reminiscence mentioned above did not occur for the starting latencies. (*b*) The superiority of 50% reinforcement to 100% in late acquisition within the 128-trial condition was even less significant than with the running speed measure (t of .91 for the last 4 trials).

Extinction.—As in acquisition, the results for starting latency were so similar to those for running speed that only the latter are reported here. The extinction curves for the running speed measure are shown for the 5 100% groups in Fig. 2 and for the 5 50% groups in Fig. 3. All 10 groups were combined in a trend analysis, in which the four trend points were, respectively, the last 4 trials of acquisition, the first 12 trials of extinction, the next 16 trials of extinction, and the final 12 trials (second day) of extinction. The acquisition point was included in order to take account of extinction occurring on the first few trials. This analysis is summarized in Table 1. Probability values remain the same if the conservative degrees of freedom suggested by Geisser and Greenhouse (1958) are substituted for those indicated. The significant *F* values indicate that groups receiving more training showed greater overall speed but also a more rapid decrease in speed over trials. The familiar partial reinforcement effect was present, but it was not very powerful, and did not enter into any interactions. The smallness of the PRE with our small reward is consistent with the finding by Hulse (1958) and by Wagner (1961) that the PRE is greater with large reward than with small.

TABLE 1

TREND ANALYSIS OF LATE ACQUISITION AND EXTINCTION

Source	*df*	*MS*	*F*
Uncorrelated			
Acquisition levels (A)	4	21.883	12.58***
Percentages (P)	1	7.308	4.20*
A × P	4	2.238	1.29
Ss within groups	90	1.740	
Correlated			
Trend (T)	3	2.986	12.98***
T × A	12	1.518	6.60***
T × P	3	.376	1.63
T × A × P	12	.397	1.73
Error	270	.230	

Note.—Meaning of trend points is explained in text.
* $p < .05$.
*** $p < .001$.

It appears from Fig. 2 and 3 that little extinction occurred in the groups with the least acquisition training. Groups with more training started at a higher level, dropped more rapidly, and in the 100% condition appear to have reached nearly a common asymptote. This is indicated by the *F* for number of acquisition trials within the 100% condition on Trials 13–28 of extinction, which is less than 1. Al-

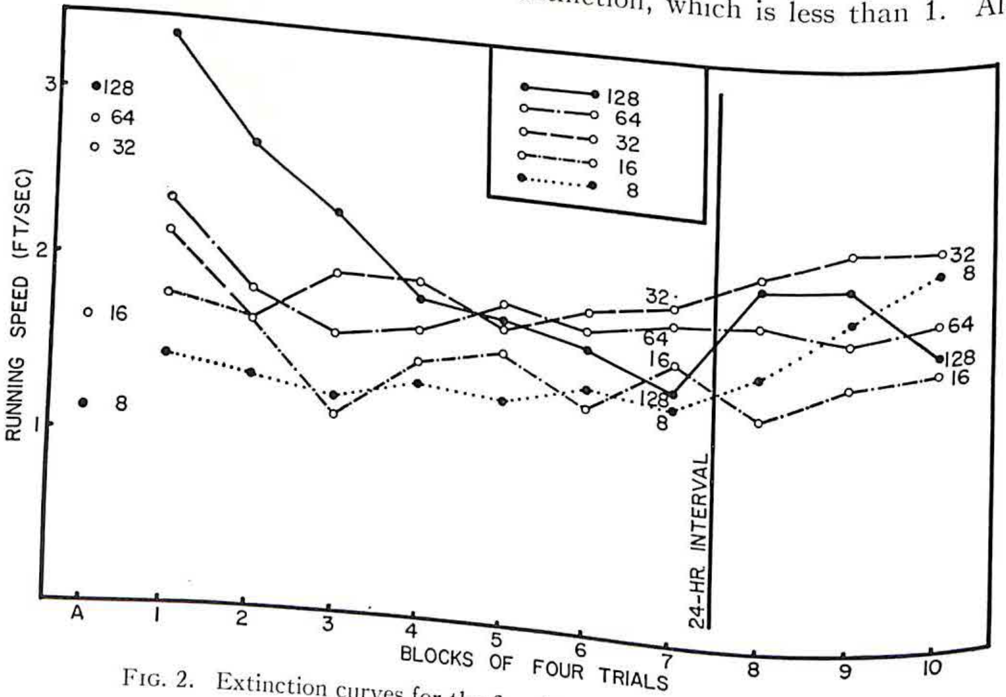


FIG. 2. Extinction curves for the five 100% groups. (A designates the final four trials of acquisition.)

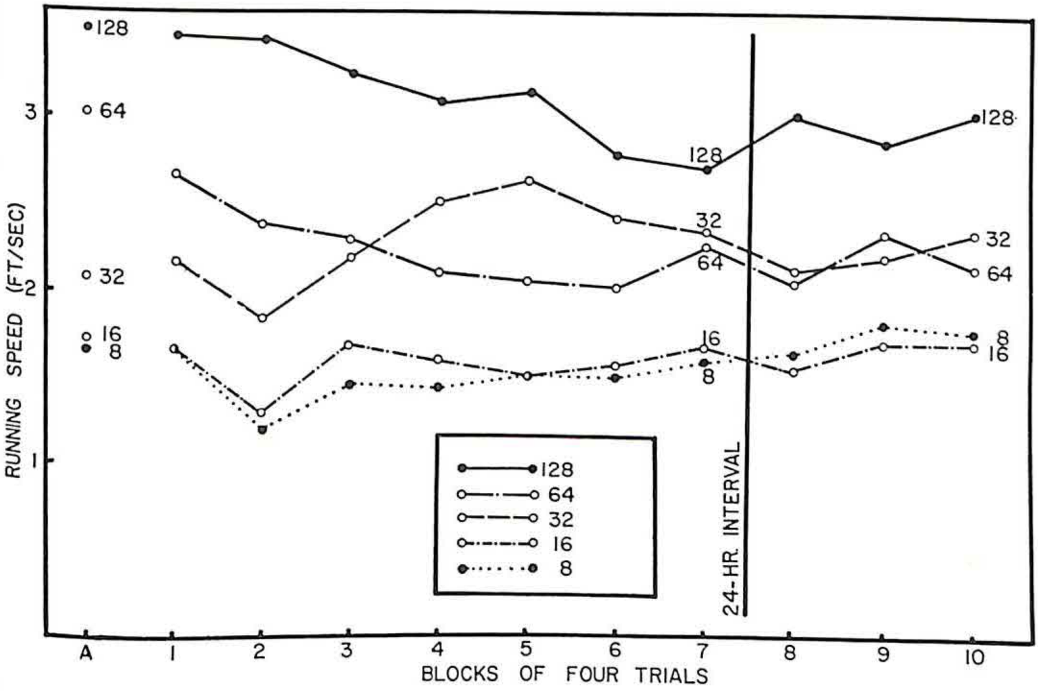


FIG. 3. Extinction curves for the five 50% groups. (A designates the final four trials of acquisition.)

though a curvilinear relationship is suggested within the 100% groups on the last four trials of the first extinction day, neither the quadratic ($F = 2.84$) nor the linear ($F < 1$) components are significant with $df = 1/45$. In the 50% condition the initial differences were preserved to a greater extent, the overall F for Trials 13–28 being 4.73, significant at the .01 level for $df = 4/45$.

DISCUSSION

The findings in this experiment, though not decisive, tend to support the view that resistance to extinction is an increasing function of number of acquisition trials throughout the range of values tested. This relationship is fairly clear-cut in the 50% condition, but in the 100% condition the conclusion depends partly on how one defines resistance to extinction. Groups with more training ran faster in early extinction, showed no difference in later extinction, and therefore dropped more rapidly in absolute terms (as revealed by the significant Trend × Acquisition Level interaction) but not necessarily in relative terms. In any case, there is no indication that additional training trials lead to reduced running speed during extinction.

The reason for the marked difference between our results and those of North and Stimmel (1960) is not clear. The most obvious difference between the two procedures is their relative spacing of trials (20–40 min. ITI) in contrast to our massing. Their rats received a larger daily ration than ours, and the reward magnitudes may also differ, but they do not specify the weight of their reward pellets. Our apparatus also differed considerably from theirs although their primary response measure of starting time was very similar to our starting time and in our experiment starting time (though somewhat more variable) yielded the same relationships as running time. The fact that our response measure revealed less extinction may be due to the width and height of our runway which were very likely more conducive to running than to competing responses. It is noteworthy that the two experiments they report differ in the detailed results, with the maximum superiority of the least overtrained group coming early in extinction in Exp. I but late in extinction in Exp. II. If the trend of our 128-trial 100% group in the first day of extinction had continued (see Fig. 2), our curves would have shown some (though still not close) resemblance to the comparable curves in North and Stimmel's Exp. II. However, the apparent spontaneous recovery in our 128-trial 100% group interrupted this trend, and our results are clearly discrepant from North and Stimmel's.

A possible interpretation of the relation found here between number of training trials and extinction is based on Spence's (1956) equation, $E = H(D + K)$, in which H depends on the number of prior trials, whether rewarded or not. At the end of acquisition, the groups differed in H but were equal in K, since K reaches its asymptotic value much more rapidly than does H (an assumption based on Hull, 1952). Early in extinction, K dropped rapidly to zero. The higher the H, the greater the resulting absolute decrement in E and hence in speed. Speeds did not return to their operant level, since H was higher in extinction than at the beginning of acquisition. The 8-trial and 16-trial groups showed no evidence of extinction because H was continuing to increase fast enough to compensate for the decrease in K. Near the end of extinction, speeds reached a common asymptote since, at this point, all groups had experienced enough trials to attain asymptotic H. (See also North & Stimmel, 1960.) This interpretation works well for the 100% groups, but additional assumptions would be needed to explain the performance of the 50% groups.

REFERENCES

CAPALDI, E. J. The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, **51**, 367–371.

EDWARDS, A. L. *Experimental design in psychological research.* (Rev. ed.) New York: Holt, Rinehart, & Winston, 1960.

GEISSER, S., & GREENHOUSE, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. *Ann. math. Statist.*, 1958, **29**, 885–891.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

HARRIS, P., & NYGAARD, J. E. Resistance to extinction and number of reinforcements. *Psychol. Rep.*, 1961, **8**, 233–234.

HILL, W. F., & SPEAR, N. E. Resistance to extinction as a joint function of reward magnitude and the spacing of extinction trials. *J. exp. Psychol.*, 1962, **64**, 636–639.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

HULSE, S. H. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

LAWRENCE, D. H., & MILLER, N. E. A positive relationship between reinforcement and resistance to extinction produced by removing a source of confusion from a technique that had produced opposite results. *J. exp. Psychol.*, 1947, **37**, 494–509.

LEWIS, D. J., & COTTON, J. W. The effect of intertrial interval and number of acquisition trials with partial reinforcement on performance. *J. comp. physiol. Psychol.*, 1959, **52**, 598–601.

LEWIS, D. J., & DUNCAN, C. P. Expectation and resistance to extinction of a lever-pulling response as a function of percentage of reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, **55**, 121–128.

MURILLO, N. R., & CAPALDI, E. J. The role of overlearning trials in determining resistance to extinction. *J. exp. Psychol.*, 1961, **61**, 345–349.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242.

(Received July 13, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 501–506

SEARCH-DISCRIMINATION TIME FOR MISSING STIMULUS INFORMATION[1]

ALEY THOMAS AND CHARLES M. SOLLEY

Center for the Study of Cognitive Processes, Wayne State University

2 experiments in search discrimination were conducted. In both, 6th grade children had to search for a missing stimulus target which would complete the set. Targets were defined by either single classifications—color, form, or size; double classifications—color-form, color-size, or form-size; and triple classifications—color-form-size characteristics. The number of stimuli in the search field was either 8, 17, 35, or 53. It was found that search-discrimination time increased linearly with the number of stimuli; that single- and triple-classified stimuli were easier than double-classified ones; and that practice facilitated search discrimination, particularly for double-classified stimuli.

In nearly all complex discrimination tasks a person must search for stimulus information and, simultaneously, make simple decisions about the stimuli he samples in his search. The present studies are concerned with the search-discrimination-inference behavior of a child when he is confronted with a number of stimuli of various classifications, of which one item is missing.

There are several variables which can make a search-discrimination-decision problem easy or difficult. The present studies used the following variables: (*a*) The number of stimuli in the perceptual field, (*b*) the number of classifications defining a search target, and (*c*) the amount of practice *S* was given at discriminating and categorizing the stimuli. Eriksen (1952) has demonstrated that an increase in the number of classifications defining search targets, produces an increase in search-discrimination time. Accordingly, it was hypothesized that as the number of classifications defining a missing stimulus target increased search-discrimination time would increase. It is also known (Archer, 1954; Gregg, 1954) that discrimination reaction time decreases with practice. Again, it was hypothesized that search-discrimination time for locating missing stimulus targets would decrease with practice. As to a hypothesis about the effects of number of stimuli, there were two alternative hypotheses, both based on experimental evidence. Several studies (e.g., Archer, 1954; Archer, Bourne, & Brown, 1955; Gregg, 1954) have shown a logarithmic relationship between the discrimination time and the number of stimuli in a field, while a few other studies (e.g., Solley & Snyder, 1958) have found a linear relationship in search-discrimination tasks. It was hypothesized, on this basis, that a monotonic relationship would be found between search-discrimination time and the number of stimuli to be searched among.

METHOD

Stimulus materials.—The stimulus materials were geometric figures cut out of colored paper and mounted at randomly selected points on sheets of white paper. They were of three colors (red, green, and blue), three forms (circle, square, and diamond), and three sizes (900 sq. mm. or large, 625 sq. mm. or medium, and 400 sq. mm. or small). A total of 27 combinations was possible if all three ways of classifying were used, but only a certain number was selected for each problem. In each of the problems one stimulus was missing. In three problems the missing stimulus was defined by a single classification, either color, form, or size. For example, in one problem there were three small red circles, three small green circles, and two small blue circles. The *S* had to search and discriminate until he could say, "the blue one is missing." A single classification specified the missing search target. In three other problems the missing stimulus had a double classification, either form-color, form-size, or color-size. For example, in one of these problems there were three blue diamonds with one of each size, three red diamonds with one of each size, and two green diamonds with one large and one medium. The *S* had to search and discriminate until he could say, "the small green one is missing." In another problem, the missing stimulus had a triple classification. For example, in this problem there were three green circles with one of each size, three red diamonds with one of each size, and two blue squares with one medium and one small. The *S* had to specify the missing stimulus, a large blue square. All of these examples were from the group of seven problems in which eight stimuli were present. Other problems were constructed using 17, 35, and 53 stimuli. For example, in a problem using 53 stimuli, there were 18 small red circles, 18 small green circles, and 17 small blue circles with *S* having to locate the missing "blue" stimulus.

---

[1] This research was partially supported by the Menninger Foundation and by a research grant (M-3924) from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service.

Subjects.—Problems with 8, 17, and 53 figures, respectively, were given to four groups of 10 *S*s. Each of these groups received one level. These groups are referred to as Exp. I. A fifth group of 10 *S*s received all of the problems at three levels (8, 17, and 35 figures). Problems with 53 figures were omitted since it was found, in preliminary testing, that all four levels of problems made the task too long and tedious for the *S*s. This fifth group, with repeated measurements over levels, is referred to as Exp. II.

The *S*s were 50 children (28 boys and 22 girls) of the sixth grade of Isaac Crary Elementary School in Detroit, Michigan. The average age of *S*s was 11 yr. Each *S* was randomly assigned to one of the five groups.

Procedure.—Each *S* was tested individually. He was shown an example of each type of stimulus which he identified. This was done to make sure the child could label the colors, the forms, and the sizes correctly. If the child called the square a box or the circle a ball, it was considered appropriate. He was then told:

> This is a game I want you to play. It is an unusual game because you have to find something that isn't there. For example, if I were to show you a picture of a car with only three wheels, you would say, "a wheel is missing." You are to look at each of these pictures and tell me what is missing. Try to find what is missing as soon as you can. In this problem there will be —— figures for you to look at. Do you understand what you are to do?

Following these instructions, *E* answered all questions until *S* said he knew what he was to do. Then *E* exposed one of the problems by placing it in front of *S*. At the same time *E* started his stop watch. If *S* gave a wrong answer, *E* told him to continue and permitted the stop watch to continue. A different random order of presenting problems was used for each *S*.

RESULTS

In order to evaluate the results on the search-discrimination experiments it is useful to examine the two studies separately. Experiment I can tell us whether or not the number of stimuli in the search field (referred to as "levels") and the type of stimulus classification influenced the time required to find what was missing. Experiment II gives us data which rechecks these variables and which gives us information, in addition, on the effects of practice.

Experiment I.—The mean search-discrimination times of *S*s, required for each situation in Exp. I, are given in Table 1.

It appears from the data that the search-discrimination time increases with the increase in the number of stimuli though the increase is not the same for all the factors. Table 1 shows that the increase is approximately linear for all types of stimulus classifications except for the color-size

TABLE 1

MEAN SEARCH-DISCRIMINATION TIME (SEC.) OBTAINED IN EXP. I

Factors	Number of Figures			
	8	17	35	53
Color	7.5	22.5	24.2	55.1
Form	6.0	14.7	22.5	47.2
Size	18.3	26.8	52.3	65.5
Color and form	16.2	24.4	28.1	67.1
Form and size	18.6	54.3	67.3	92.5
Color and size	9.7	30.8	80.3	65.0
Color, form, and size	6.6	18.4	31.0	55.4
Mean	11.9	22.4	43.7	64.0

classification. For example, the search-discrimination time for color-size for 35 figures is greater than for the level of 53 figures. In the remaining situations the increase in discrimination increases with the number of stimuli. These data indicate a linear relationship between the number of stimuli and the average time required to find the missing stimulus for each type of stimulus classification. The mean search-discrimination times, averaging over the different stimulus classifications, for each level of 8, 17, 35, and 53 figures are 11.9, 22.4, 43.7, and 64.0 sec., respectively.

As for classifications of search targets, the overall search-discrimination times for color and form are not significantly different from one another, as tested by a *t* test, while both of these are significantly lower than for a missing size. Among the double classifications, form-size is significantly harder to locate than color-form but none of the other differences is significant. The triple classification, color-form-size, is significantly easier to locate than any of the double classifications but is not significantly different from the single classifications. These tests, however, did not bear directly upon our hypothesis concerning the effects of the number of classifications defining a search target.

To test this hypothesis, the average search-discrimination times for single, double, and triple classifications of missing targets were computed for each individual. The analysis of these data is given in Table 2.

The *F* between levels is significant, indicating that the number of stimuli in the search field significantly affects search time. In addition, there is a significant difference between number of classifications defining the search target. As determined by the *t* tests, the double-classification problems are significantly more difficult than the single- or triple-classification problems. The significant interaction of Levels × Dimensions indicates that the slopes shown in Fig. 1 are not parallel. The slope for the two-factor problems is steeper than those for the single- or triple-classification problems.

The mean search times for single, double, and triple classifications of stimuli were plotted against number of stimuli, and a linear function was fitted by the method of least squares. The results are shown in Fig. 1. This figure shows that a linear relationship is much more reasonable than a logarithmic one. It also shows what

TABLE 2

ANALYSIS OF VARIANCE OF SEARCH-DISCRIMINATION TIMES IN EXP. I

Source	*df*	*MS*	*F*
Uncorrelated			
Levels (L)	3	161,055.5	17.2**
Error (b)	36	936.0	
Correlated			
Classifications (C)	2	31,074.5	203.2**
L × C	6	2,516.9	16.5**
Error (w)	72	152.9	
Total	119		

** $p < .01$.

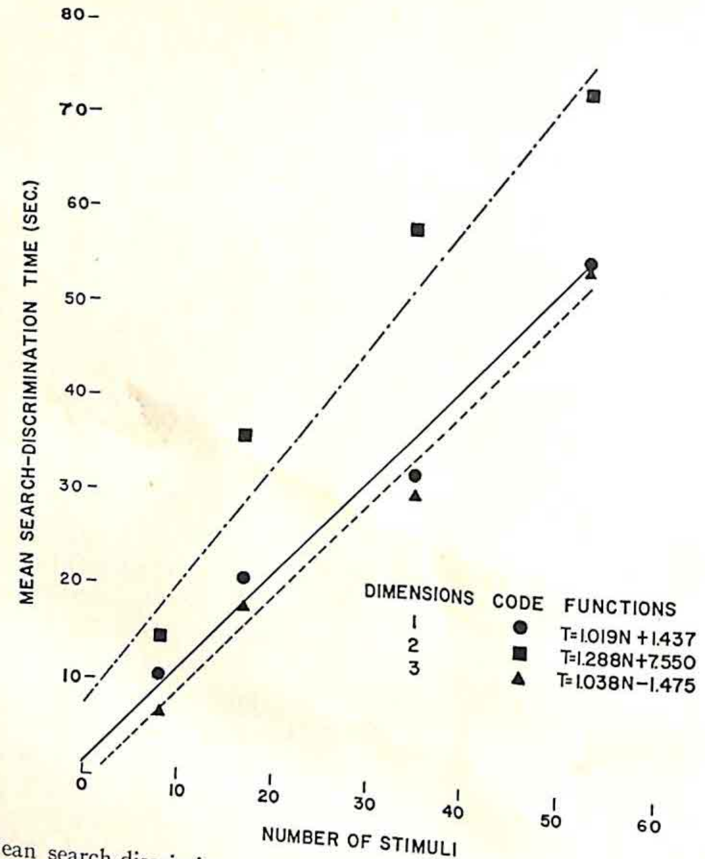


FIG. 1. Mean search-discrimination time as a linear function of number of stimuli in search fields composed of single-, double-, and triple-classified stimuli. (Least squares solutions are shown for data from Exp. I.)

our *t* tests on specific factors indicated, namely, that there is no significant difference between times for single- and triple-classified targets and that double classifications are much more difficult to locate.

Experiment II.—An analysis of variance was run on the data for Exp.

TABLE 3

ANALYSIS OF VARIANCE OF SEARCH-DISCRIMINATION TIME IN EXP. II

Source	*df*	*MS*	*F*
Ss	9	755.1	
Classifications (C)	6	634.8	5.8**
Ss × C	54	131.0	
Levels (L)	2	9,105.2	24.9**
Ss × L	18	369.0	
C × L	12	97.6	2.2*
Ss × C × L	108	43.8	

* $p < .05$.
** $p < .01$.

TABLE 4

MEAN SEARCH-DISCRIMINATION TIME (SEC.) OBTAINED IN EXP. II

Factors	Number of Figures		
	8	17	35
Color	7.7	8.9	20.2
Form	7.7	15.8	24.8
Size	11.6	19.0	36.6
Color and form	10.7	16.7	39.2
Form and size	15.2	21.2	48.8
Color and size	5.8	15.6	34.1
Color, form, and size	5.0	12.9	31.7
Mean	9.1	15.7	33.6

II. Since each *S* had a score in each condition, a mixed design was used. The results of this analysis are shown in Table 3. This shows that all variables were significant.

The means for the various experimental conditions in Exp. II are shown in Table 4. As in Exp. I, *t* tests reveal no significant difference between color and form but that each of these single classifications was easier than size. There were no significant differences among the double-classified targets although form-size was again the most difficult to locate. Form-size was more difficult than color alone, form alone, or color-form-size. Again, there was no difference between color-form-size and color alone, or form alone.

As in Exp. I, a mean score for single-classified, double-classified, and triple-classified search targets was computed for each *S*. Again, *t* tests reveal that there were no significant differences between single and triple classifica-

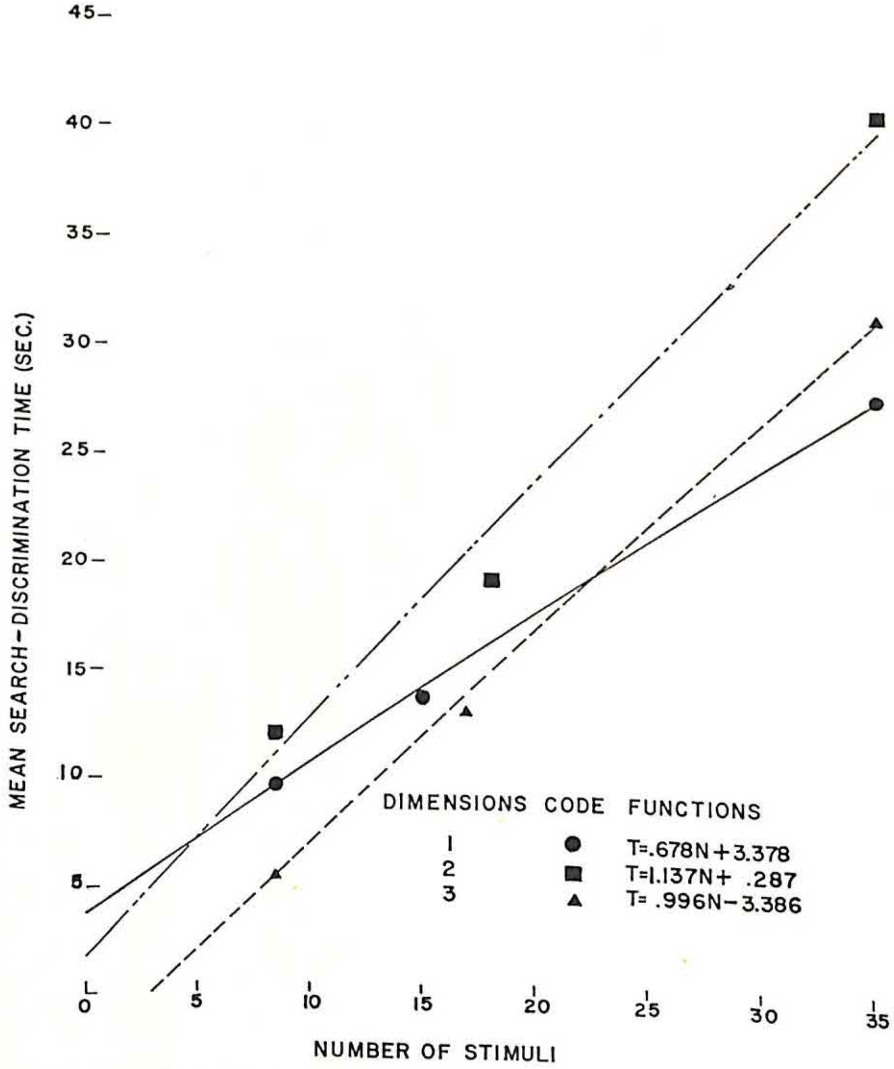


FIG. 2. Mean search-discrimination time as a linear function of number of stimuli in search field composed of single-, double-, and triple-classified stimuli. (Least squares solutions are shown for data from Exp. II.)

tions, and that both of these were significantly easier than double classifications.

The mean search-discrimination times for single-, double-, and triple-classified search targets were plotted as a function of number of stimuli to be searched among, as shown in Fig. 2. A linear line again seemed to yield the best fit, and the least squares solutions are shown in Fig. 2.

Since all 3 levels of problems were given to the same group of students in Exp. II and since these *S*s were a random sample from the same population as in Exp. I, it was possible to note the effects of practice. In this, the mean discrimination times for the 8 levels for Exp. I and II were not significantly different but the times for the 17 and 35 levels were, showing that practice did facilitate search. A closer inspection of Tables 1 and 4 reveals that only the double-classification problems showed a substantial reduction in search time as a function of practice; and that most of this reduction came for form-size and color-size.

Discussion

Contrary to what information theory would predict, a logarithmic relationship between search-discrimination time and number of stimuli in the search field was not found. Instead, a simple linear relationship was found similar to that reported by Solley and Snyder (1958). This is difficult to rationalize on an a priori basis but stands as an empirical finding.

In agreement with Eriksen (1952) and Archer (1954) are the results showing that double-classification problems were more difficult than are single-classification ones. However, the triple-classification problems were not more difficult than the double; rather, the reverse was true. Indeed, the triple-classified search targets were as easy to find as the single-classified ones. This result was completely unexpected but held up from level to level and from Exp. I to Exp. II.

In addition, a comparison of Exp. I with Exp. II revealed that practice facilitated search discrimination, especially with respect to search for double-classified targets. The *S*s appeared to be more at ease with the task and it was suspected that practice facilitated understanding the task requirements.

References

Archer, E. J. Identification of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification of irrelevant information and instruction. *J. exp. Psychol.*, 1955, **49**, 159–164.

Eriksen, C. W. Location of objects in a visual display as a function of the number of dimensions along which the objects differ. *J. exp. Psychol.*, 1952, **44**, 56–60.

Gregg, L. W. The effect of stimulus complexity on discrimination responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

Solley, C. M., & Snyder, F. W. Information processing and problem solving. *J. exp. Psychol.*, 1958, **55**, 384–387.

(Received July 14, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 507–514

TRANSFER FROM VERBAL-DISCRIMINATION TO PAIRED-ASSOCIATE LEARNING: II. EFFECTS OF INTRALIST SIMILARITY, METHOD, AND PERCENTAGE OCCURRENCE OF RESPONSE MEMBERS[1]

WILLIAM F. BATTIG AND H. RAY BRACKETT

University of Virginia

In agreement with previous results, paired-associate (PA) performance was not significantly facilitated on pairs identical to those of a previous verbal-discrimination (VD) list, whereas poorer performance resulted if the VD and PA items were paired differently, as compared with PA items excluded from the VD list. These transfer effects were unaffected by intralist similarity or percentage occurrence of response members (ORM), but were largely eliminated under the recall method of PA learning. Contrary to previous findings, overall recall-method PA performance was only insignificantly superior to that under the standard anticipation method. Moreover, a contrasting distinct superiority of the anticipation method was found under conditions of 50% ORM, especially for medium intralist similarity.

In a previous study, prior verbal-discrimination (VD) learning did not facilitate subsequent paired-associate (PA) learning, even when exactly the same pairs occurred in both the VD and PA lists (Battig, Williams, & Williams, 1962). This study also showed significant interference with PA performance if items were paired differently in the VD and PA lists. In addition, the present authors have reported PA learning of pairs of nonsense shapes and two-digit numbers under the standard *anticipation* method to be significantly inferior to a *recall* method, wherein each trial begins with the presentation of the entire list of pairs for learning followed (in a different order) by the entire set of stimulus (S) terms alone for a recall test (Battig & Brackett, 1961).

Intralist similarity conditions were suggested as a possible determinant of both of these largely unexpected findings. Specifically, the absence of positive transfer from VD to PA learning may have resulted from the use of CVCs of low similarity (Battig et al., 1962). Likewise, the level of similarity (or other factors) inherent in the shape-number pairs may account for the superiority of the recall over the anticipation method (Battig & Brackett, 1961; see also Lockhead, 1962). An alternative explanation for the latter finding may reside in the fact that, within each recall-method trial, presentation of the entire list of pairs for learning precedes and is completely separated temporally from the subsequent presentation of S terms alone for performance measurement. The temporal confounding of these learning and performance-measurement functions which characterizes the anticipation method may therefore be sufficient to override the presumed advantages therein of immediate knowledge of results produced by the presentation of each S-R pair following its S term (Battig & Brackett, 1961, pp. 63–64). If so, partial (50%) rather than 100% occurrence of response members (ORM) on each trial should eliminate this separation of functions under the recall method and reduce or eliminate its superiority to the anticipation method. Since 50% ORM conditions should also increase the relative importance of S term learning, to which VD learning has been indicated to be more closely related than to R term learning (Battig et al., 1962), more transfer from VD learning might be expected to PA learning under 50% than 100% ORM conditions. Consequently, the present study was designed to investigate possible differential effects of intralist similarity and %ORM on transfer from prior VD learning to PA learning under recall and anticipation methods.

---

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare. Preparation of the manuscript was supported by funds from Grant M-5769 from the National Institute of Mental Health, United States Public Health Service.

METHOD

Subjects.—The 144 *S*s were all summer session students or affiliates of the University of Virginia who were assigned to the various experimental conditions on the basis of a predetermined unsystematic sequence (based on order of appearance in the laboratory) which provided that the number of *S*s in each condition remained equal at all stages of the experiment. An additional 6 *S*s were eliminated, on the basis either of failure to follow instructions or refusal to continue with the experiment, and were replaced by the next *S* to be run.

Procedure.—Each *S* first learned a verbal-discrimination (VD) list (i.e., to pronounce the arbitrarily selected "right" syllable of each of 12 CVC pairs presented under self-paced conditions in unsystematically varying serial and spatial order from trial to trial), following the identical procedure described in detail previously (Battig et al., 1962, p. 259). Immediately after the first errorless VD trial, each of the 24 CVCs was presented individually in an unsystematic order, while *S* attempted to recall the CVC previously paired with it in the VD list. This VD recall test was designed as a check on the degree to which intrapair associations had been learned during the VD task. To minimize the possibility of additional learning of intrapair associations during VD recall, *S* was given no information as to the correctness of his responses.

Each *S* was then taken to another part of the experimental room, previously blocked from his view, and seated in front of a projection screen. The 16-mm. film containing the appropriate PA list in four different serial

TABLE 1

PA AND SAMPLE VD LIST FOR EACH SIMILARITY CONDITION

VD List			PA List			Type of Pair	Mean CRs
High Sim	Med Sim	Low Sim	High Sim	Med Sim	Low Sim		
R W[a]	R W	R W	S R[b]	S R	S R		
MER-LIB	GAR-TIF	HAK-DEX	MER-LIB	GAR-TIF	HAK-DEX	Disc(S)	12.28
MIR-FAL	KIP-FEB	LEV-SIM	MIR-FAL	KIP-FEB	LEV-SIM	Disc(S)	
RAF-KUM	TER-CAV	HUP-COL	KUM-RAF	CAV-TER	COL-HUP	Disc(R)	13.09
LOR-KUS	LOR-KUS	LOR-KUS	KUS-LOR	KUS-LOR	KUS-LOR	Disc(R)	
MUS-mos	MUC-mov	MUC-buk	MUS-BIR	MUC-REN	MUC-GAT	Fam(S-R)	11.01
BIR-ber	REN-fak	GAT-gul	SIM-FER	VEG-LOK	ZOR-JAS	Fam(O)	12.17
rul-SIM	ger-VEG	ZOW-ZOR	KUR-FOB	CIR-GAT	FET-WUN	Fam(S)	11.81
feb-FER	pok-LOK	pax-JAS	MIK-RAB	GIM-NEK	JIB-KOY	Fam(R)	12.09
KUR-mur	CIR-cis	FET-yeh	RES-BOL	PEK-GIB	NOX-TIZ	New	12.34
ruf-FOB	tor-GAT	vin-WUN	SEM-RIF	SEP-KAL	VAG-PUD	New	
muk-MIK	gol-GIM	tec-JIB	SIK-LAR	VAS-NOG	YAW-FEB	New	
RAB-rok	NEK-nub	KOY-dop	REM-BIF	REM-BIF	REM-BIF	New	

[a] R = right; W = wrong.
[b] S = stimulus; R = response.

orders was presented by a Dunning animatic strip-film projector. Learning was to a criterion of 2 successive errorless trials or a maximum of 24 trials.

As illustrated in Table 1, the 12 PA pairs for all *S*s included four pairs representing each of three basic types of relationships to the pairings of the VD task. The discriminated (Disc) pairs, identical in the two lists, included two for which the S term was "right" and the R term "wrong" on the VD task, designated as Disc(S) pairs, whereas for the two Disc(R) pairs the S term had been wrong and the R term right. The four familiarized (Fam) pairs were constituted from CVCs which had been paired in the VD list with other CVCs not appearing in the PA list (designated in lower case letters in Table 1). The Fam pairs were subdivided so as to represent all possible combinations of number and position of right and wrong VD syllables, designated as Fam(S-R), Fam(S), Fam(R), and Fam(O) pairs (the symbols in parentheses designating right VD syllables). The remaining four PA pairs were New pairs composed of syllables not appearing in the VD list.

Groups.—Twelve equal sized groups represented all possible combinations of three conditions of intralist similarity, two methods (recall and anticipation), and two %ORM conditions (100% and 50%) as described more fully below.

Intralist similarity.—All *S*s in the high similarity (High Sim) groups learned the same PA list consisting of S terms constructed from four consonants (KMRS) and three vowels (EIU), with R terms likewise comprised of four consonants (BFRL) with the vowels A, I, and O (except for one E). The single PA list for all medium similarity (Med Sim) groups was composed of S terms representing eight consonants (CGKMPRSV) and four vowels (AEIU), and of R terms also consisting of eight consonants (BFGKLNRT) and four vowels (AEIO). The low similarity (Low Sim) list used all consonants (except Q) and all vowels with approximately equal frequency, but no consonant was replicated as the first or last letter within either the set of S or R terms. Syllables of all three lists, which are presented in Table 1, were of 75–85% Archer (1960) association value, and were paired so as to minimize pre-existing intrapair associations and to avoid repetition of any letter within a given pair. Partial overlap among the letters composing the S and R term subsets was approximately comparable between the three lists. Of the eight additional CVCs appearing in the VD list but not in the PA task, four had the same letter composition as the S terms and four corresponded to the R terms of the PA list within each similarity condition.

Recall-anticipation method.—Under the recall method of PA learning, each trial began with the successive presentation of the 12 S-R pairs for 2 sec. each. This was followed immediately by the presentation of the 12 S terms alone in a different order, also for 2 sec. each, while *S* attempted to pronounce the correct associated R term (but was not told whether or not he was correct). Under the anticipation method, corresponding closely to that of the previous study (Battig et al., 1962), each S term was presented first for 2 sec. and was followed immediately by its S-R pair for an additional 2 sec. However, the first anticipation-method trial was preceded by the presentation of the entire list of S-R pairs for 2 sec. each, exactly as in the first half of each recall-method trial, in order to make the two methods directly comparable in terms of events preceding measurement of first-trial performance (Battig & Brackett, 1961). A 1-min. interval followed this initial pair-presentation series for both methods, to allow *E* to change the strip film under the anticipation method, and an 8-sec. intertrial interval was used thereafter.

Percentage occurrence of response members (%ORM).—In addition to the usual 100% ORM procedure whereby each S-R pair is presented on each trial, a partial (50%) ORM procedure was employed, in which half of the R terms were omitted from the normal S-R pair presentation during each trial. Under these 50% ORM conditions, a prearranged sequence assured that each R term occurred only once in each pair of trials, without being repeated or omitted on more than two successive trials. However, all S-R pairs were presented on the first recall-method trial and in the initial pair-presentation series preceding the first anticipation-method trial. To make performance measures comparable to the 100% ORM condition, responses given during what normally would have been S-R pair presentations with R terms omitted were not counted for the 50% ORM groups.

All *S*s in all groups under a given similarity condition learned exactly the same PA list. However, 12 different VD lists were used for the 12 *S*s in each group, so that each of the 12 PA pairs represented equally often the Disc, Fam, and New subsets, as well as the various S-R right-wrong combinations within the Disc and Fam subsets. A sample one of these 12 VD lists for each similarity condition has been included in Table 1.

RESULTS

Table 2 presents mean total correct responses per PA pair, on Disc, Fam, and New subsets and on the three subsets combined, for each of the 12 groups as well as for each of the three between-group variables and combinations thereof. As in the previous study (Battig et al., 1962), separate tabulations and analyses were performed on number of errors (trials) preceding the first correct response to each pair, and on percentage of trials after the first correct response on which errors were made. Since these before- and aftererrors measures both revealed essentially the same picture as the correct-responses measure, although usually at lesser levels of statistical significance, the means have not been presented and will be discussed only where they showed substantially different effects as compared with the correct-responses measure.

TABLE 2

MEAN CORRECT RESPONSES PER PAIR FOR GROUPS AND CONDITIONS

Method—ORM Condition	Type of Pair	Similarity Condition			
		High Sim	Med Sim	Low Sim	Total
Recall—100%	Disc	6.75	16.15	19.02	13.97
	Fam	8.85	17.25	17.35	14.49
	New	7.06	16.08	18.58	13.91
	Total	7.56	16.49	18.32	14.12
Recall—50%	Disc	9.04	7.77	15.29	10.70
	Fam	8.40	7.85	12.81	9.69
	New	8.23	9.10	15.42	10.92
	Total	8.56	8.24	14.51	10.44
Recall—Total	Disc	7.89	11.96	17.16	12.34
	Fam	8.63	12.55	15.08	12.09
	New	7.65	12.59	17.00	12.41
	Total	8.06	12.37	16.41	12.28
Anticipation—100%	Disc	10.17	14.10	16.85	13.71
	Fam	7.56	12.75	14.94	11.75
	New	8.52	13.21	16.35	12.69
	Total	8.75	13.35	16.05	12.71
Anticipation—50%	Disc	6.73	14.77	15.54	12.35
	Fam	6.77	12.83	13.83	11.15
	New	7.35	13.73	14.44	11.84
	Total	6.95	13.78	14.60	11.78
Anticipation—Total	Disc	8.45	14.44	16.20	13.03
	Fam	7.17	12.79	14.39	11.45
	New	7.94	13.47	15.40	12.27
	Total	7.85	13.57	15.33	12.25
Total—100%	Disc	8.46	15.12	17.94	13.84
	Fam	8.21	15.00	16.15	13.12
	New	7.79	14.65	17.47	13.30
	Total	8.15	14.92	17.18	13.42
Total—50%	Disc	7.88	11.27	15.42	11.52
	Fam	7.58	10.34	13.32	10.42
	New	7.79	11.42	14.93	11.38
	Total	7.75	11.01	14.56	11.11
Total—Total	Disc	8.17	13.20	16.68	12.68
	Fam	7.90	12.67	14.73	11.77
	New	7.79	13.03	16.20	12.34
	Total	7.95	12.97	15.87	12.26

Group differences.—Table 2 indicates a sizeable overall decrease in PA correct responses with increasing intralist similarity, shown by analysis of variance to be highly significant, $F\ (2, 132) = 48.03$, $p < .001$. The overall inferiority of 50% to 100% ORM performance was likewise highly significant, $F\ (1, 132) = 12.02$, $p < .005$, but the slight and inconsistent superiority of the recall over the anticipation method fell far short of significance ($F < 1$). However, a significant Methods × %ORM interaction, $F\ (1, 132) = 4.24$, $p < .05$, suggests the recall method to be superior under 100% ORM conditions, whereas anticipation-method performance was better for the 50% ORM groups. This latter interaction may reflect, at least in part, the depressed performance of a single group (Med Sim-recall-50%-ORM) which led further to a significant third-order Similarity × Methods × %ORM interaction, $F\ (2, 132) = 6.18$, $p < .005$. Neither the interaction between similarity and methods ($F < 1$) or similarity and %ORM, $F\ (2, 132) = 2.37$, $p = .10$, proved to be significant.

Disc-Fam-New comparisons.—As shown in Table 2, overall PA performance was only slightly better on Disc than on New pairs, and poorest on Fam pairs, in substantial agreement with previous results (Battig et al., 1962). Analysis of variance revealed the overall Disc-Fam-New difference to be significant, $F\ (2, 1479) = 4.26$, $p < .025$, primarily due to the superiority of Disc over Fam pairs ($F = 8.34$, $p < .01$), whereas the difference in favor of New over Fam pairs fell short of significance ($F = 3.27$, $p < .10$) as did that for Disc over New pairs ($F = 1.17$). Although considerable discrepancies are evident in Table 2 with respect to relative Disc-Fam-New performance within the various groups, none of the Group × Disc-Fam-New interactions achieved statistical significance, and only the Disc-Fam-New × Methods interaction even approached the .10 level, $F(2, 1479) = 2.25$. However, the Disc-Fam-New difference was highly significant for anticipation-method groups analyzed separately ($F = 6.47$, $p < .01$) but virtually absent for recall-method groups ($F < 1$). It thus appears that differential transfer effects from the various types of VD pairings may be evident only in PA performance under anticipation-method conditions. The only other indication that relative Disc-Fam-New performance differed under the various conditions of the present study was provided by a significant Disc-Fam-New × Methods × %ORM interaction emerging from the before-errors analysis, $F(2, 1479) = 4.30$, $p < .025$. This interaction reflects primarily the greater number of before-errors on Fam pairs, as compared with Disc and New pairs, under anticipation-100%-ORM and recall-50%-ORM conditions, as contrasted with the virtual equivalence of the three subsets both for anticipation-50%-ORM and recall-100%-ORM groups.

VD right-wrong comparisons.—In contrast with previous findings of better PA performance on Disc(S) pairs (where the S term was right and the R term wrong on the VD task) as compared with Disc(R) pairs (Battig et al., 1962), a sizable but nonsignificant overall difference emerged in favor of Disc(R) over Disc(S) pairs under the present conditions, $F(1, 1479) = 3.26$, $p < .10$. However, in agreement with previous results, no evidence was found of systematic or consistent differences as a function of either number or position of right and wrong VD syllables in Fam pairs ($F = 1.40$). Since none of these comparisons showed any systematic differences for the various groups, the largest of the group interactions falling well short of the .10 level of significance, only the total means have been presented in the last column of Table 1.

VD recall test.—As in the previous study, comparisons were made between PA performance on Disc and Fam pairs which did and did not include items to which correct associations had been made on the recall test immediately following VD learning, designated, respectively, as *associated* and *nonassociated* pairs (these intra-*S* comparisons necessarily excluded those *S*s for whom all pairs of a given subset were either associated or nonassociated). The previous findings of significantly better performance on associated as compared with nonassociated pairs were replicated here only for the Disc pairs representing the Low Sim lists, $F(1, 34) = 5.16$, $p < .05$, but not for either the Med Sim or High Sim conditions (both *F*s < 1). Although there was a very slight overall numerical superiority in performance on associated over nonassociated Fam pairs, this difference fell far short of significance for all similarity conditions (all *F*s < 1). In agreement with the previous study, individual VD recall performance was significantly correlated with overall PA performance ($r = .311$), a relationship which was reduced somewhat ($r = .233$) when the even closer group-level relationship between the two measures ($r = .561$) was partialed out. However, of the between-group variables, only similarity was significantly related to VD recall performance, $F\ (2, 132) = 3.36$, $p < .05$, due to substantially higher VD recall scores on Low Sim lists as compared with Med Sim and High Sim lists, which proved to be virtually equivalent in this respect.

VD performance.—As in the previous study, both errors ($r = -.300$) and trials ($r = -.312$) to VD criterion were moderately correlated with correct PA responses. However, despite the substantial differences in similarity of VD lists for the three conditions of intralist similarity, differences of borderline significance in VD performance were obtained only for the linear component of similarity for the VD errors measure, $F\ (1, 132) = 4.00$, $p < .05$. Groups differing in method and %ORM were well matched in terms of VD performance (all Fs < 1). Performance on the VD recall test was found not to be related either to VD errors ($r = .08$) or trials ($r = .07$) confirming the results of the previous study.

Discussion

The overall differences in PA performance on Disc, Fam, and New pairs obtained herein are strikingly similar to those reported previously (Battig et al., 1962), thereby significantly extending the generality of findings of no positive transfer from identical pairings on VD and PA lists, and of negative transfer for Fam pairs, where the VD and PA pairings differ. Only the slight inferiority of New- to Fam-pair performance for the High Sim groups offers even suggestive support for the initial speculation that increasing positive transfer from VD to PA learning would result at higher levels of intralist similarity, and the overall Disc-Fam-New × Similarity interaction fell far short of significance. Transfer for Disc and Fam pairs was likewise unaffected by the %ORM variable, but appeared to have been substantially reduced or eliminated under the recall as compared with the anticipation method, even though the Disc-Fam-New × Methods interaction failed to reach statistical significance. While there were some inconsistencies among the various groups as to the direction and magnitude of Disc-Fam-New differences, these were neither systematic nor statistically significant, and therefore cannot legitimately be regarded as anything other than sampling errors within the various individual groups.

There are, however, two major discrepancies between the present and previous results concerning VD-PA transfer. One of these is the present finding of near significant superiority of performance on Disc(R) as compared with Disc(S) pairs. Since the opposite superiority of Disc(S) pairs had proved to be significant in only one of the previous two experiments (Battig et al., 1962), and in neither case was there any effect of number or position of right and wrong VD syllables of Fam pairs, it would now appear that this variable has little if any consistent effect on PA performance. Coupled with the failure to find greater positive transfer effects under those conditions (recall method and 50% ORM) previously hypothesized to increase relative amount of S term learning, these results offer little support for the previous conclusion that the learning processes involved in VD learning are most closely related to S term learning in the PA task.

Secondly, PA performance in the present study was not better on associated Fam pairs (those containing CVCs which had been correctly associated on the VD recall test) than on nonassociated Fam pairs, as had been the case previously. Moreover, although the previously obtained significant superiority of PA performance on associated over nonassociated Disc pairs was replicated under the present Low Sim conditions, corresponding differences were not found for the Med Sim and High Sim groups. In interpreting these results in the previous study, it was suggested that the facilitation observed on associated Fam pairs was due to greater response learning or availability of their component syllables as compared with those of nonassociated pairs (Battig et al., 1962, p. 266). In line with this interpretation, the higher (75–85%) Archer (1960) association values of the present CVCs, as compared with the 34–37% association value CVCs employed in the previous study, may have reduced response learning sufficiently to eliminate the difference favoring associated over nonassociated Fam pairs in the present study. Nevertheless, since PA performance was not inferior on associated to nonassociated Fam pairs, the previous conclusion still holds that the interference with PA performance on Fam pairs cannot be explained in terms of the establishment of incorrect interfering VD associations as measured by the VD recall test.

The present results offer considerably less support for the generality of the previous finding of better PA performance under the recall as compared with the anticipation method (Battig & Brackett, 1961), although there is no evidence to support the hypothesis that this difference decreases systematically with decreasing intralist similarity. However, the significant Methods × %ORM interaction, indicating that the overall superiority of recall over anticipation groups under 100% ORM was reversed under 50% ORM conditions, is consistent with an explanation of the previously obtained superiority of the recall method in terms of its complete temporal separation of presentation of S-R pairs for learning and of S terms alone for recall. Under 50% ORM recall-method conditions, this sharp separation of learning and performance-measurement functions is necessarily reduced, thereby introducing a mixed sequence of S term and S-R pair presentations more closely comparable with that characterizing the anticipation method with either 100% or 50% ORM. Nevertheless, as shown by the even larger third-order interaction of methods and %ORM with similarity, the inferiority of recall-method performance under 50% ORM reflects primarily the markedly depressed performance of a single (Med Sim-recall-50%-ORM) group. While the performance of this deviant group on both the VD task and VD recall test was not out of line with the other groups, neither of these measures was sufficiently correlated with PA performance to argue convincingly against the possibility of accounting for the overall Methods × %ORM interaction in terms of a chance difference between the Med Sim-recall-50%-ORM group and the other groups in level of learning ability.

In any event, even if the obtained general superiority of other recall- over anticipation-method groups is to be accepted on grounds of a sampling error, the magnitude of the differences remains far below that observed previously with the shape-number pairs (Battig & Brackett, 1961). This result, along with Lockhead's (1962) finding of no significant anticipation-recall method difference with CVC materials, points to kind of material as a likely important determinant of relative performance under the two methods. In this connection, however, it should be noted that two recently completed studies in this laboratory have shown recall-method performance on low similarity CVC lists to be significantly and consistently superior to the anticipation method (*a*) over a wide range of conditions of S and R term meaningfulness; (*b*) under both constant and varied serial orders; and (*c*) for bidirectional (both S-R and R-S associations) as well as unidirectional learning conditions.

Although the present sizable differences in PA learning between similarity conditions are entirely consistent with previous results (e.g., Underwood, 1953), the failure to find clear-cut differences due to similarity on the VD task is not (Underwood & Archer, 1955), which may reflect the insensitivity of available VD performance measures under the present self-paced conditions. The overall inferiority of 50% to 100% ORM conditions also substantiates previous results (Goss, Morgan, & Golin, 1959; Goss, Nodine, Gregory, Taub, & Kennedy, 1962; Schulz & Runquist, 1960; Wilcoxon, Wilson, & Wise, 1961), but the direction of the present nonsignificant %ORM × Similarity interaction, indicating greater differences between 100% and 50% ORM for low than for high similarity lists, is directly opposed to the significant interactions for both S and R term similarity reported by Goss et al. (1962). This discrepancy may reflect the longer lists and lower learning criterion employed herein, or any of several procedural differences between the two studies. In addition, the present finding of an insignificantly greater percentage of errors after the first correct response under 50% ORM conditions more closely agrees with Wilcoxon et al. (1961), than with Schulz and Runquist (1960) who had reported fewer such 50% ORM errors.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

BATTIG, W. F., & BRACKETT, H. R. Comparison of anticipation and recall methods in paired-associate learning. *Psychol. Rep.*, 1961, **9**, 59–65.

BATTIG, W. F., WILLIAMS, J. M., & WILLIAMS, J. G. Transfer from verbal-discrimination to paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 258–268.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associate learning as a function of percentage of occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **57**, 96–104.

GOSS, A. E., NODINE, C. F., GREGORY, B. N., TAUB, H. A., & KENNEDY, K. E. Stimulus characteristics and percentage of occurrence of response members in paired-associates learning. *Psychol. Monogr.*, 1962, **76**(12, Whole No. 531).

LOCKHEAD, G. R. Methods of presenting paired associates. *J. verbal Learn. verbal Behav.*, 1962, **1**, 62–65.

SCHULZ, R. W., & RUNQUIST, W. N. Learning and retention of paired adjectives as a function of percentage occurrence of response members. *J. exp. Psychol.*, 1960, **59**, 409–413.

UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142.

UNDERWOOD, B. J., & ARCHER, E. J. Studies of distributed practice: XIV. Intralist similarity and presentation rate in verbal-discrimination learning of consonant syllables. *J. exp. Psychol.*, 1955, **50**, 120–124.

WILCOXON, H. C., WILSON, W. R., & WISE, D. A. Paired-associate learning as a function of percentage of occurrence of response members and other factors. *J. exp. Psychol.*, 1961, **61**, 283–289.

(Received July 18, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 515–519

INTERPOLATED ACTIVITY AND THE LEARNING OF A SIMPLE SKILL[1]

KENNETH A. BLICK AND EDWARD A. BILODEAU

Tulane University

2 experiments compared the effects of interpolated activity (R_b) on the acquisition of a simple arc-drawing response (R_a). The sequence of events was R_a-KR, R_b, R_a-KR, R_b, etc. For 1 group of Experiment I, R_b was identical to R_a, for 3 groups R_b was quantitatively similar, and for a final group R_b was opposite in direction. In Experiment II a rest-control group was used to evaluate the direction (sign) of transfer. The results indicated that the learning of R_a was not differentially affected by the similarity of R_b. The interpolated activity was also found to have no negative or positive transfer effects on the acquisition of R_a. The findings were related to several other null-producing procedures and their relevance to the Osgood Surface was discussed.

The evidence for interference in the learning of motor skills has never been strong. One example is that of Lewis and his 15 yr. of research using the OL-IL-RL paradigm. Another example is the OL-RL paradigm which has also produced little or no forgetting (Bilodeau & Bilodeau, 1961). The present experiment will only reassert the difficulties of producing experimental interference through retroaction.

Negative transfer effects have not been found in studies of the learning of relatively simple, discrete motor responses (Bilodeau & Bilodeau, 1958; Larré, 1961). After *S* made a response, various response events were interpolated before or after KR (knowledge of results) in attempts to modify the learning of subsequent responses. The learning of a simple positioning response was not affected. The interpolated response, however, was somewhat different from the response which *E* required *S* to learn. That is, it was usually of the wrong amplitude, duration, or effort.

In view of these findings and those of others (Annett, 1959; Bilodeau, Bilodeau, & Schumsky, 1959; Bilodeau & Ryan, 1960), the present experiment was designed with the possibility in mind that the results might be null. Nevertheless, it was believed that the procedures to be used would maximize the chances of producing nonnull results—if such procedures exist.

A relatively rare procedure in motor skills would be to control the type and amount of response which is to be interpolated. In previous studies, the interpolated response was left to *S*, and this procedure resulted in responses which closely resembled the one which followed the *last* KR. A study is now required where the response is regulated by *E*. This is suggestive of the Osgood Surface (1949) which long ago indicated that variation of interpolated responses would produce particular transfer effects. Control over responses is not easily achieved, but in the procedure below the interpolated responses were satisfactorily scaled and varied from group to group.

EXPERIMENT I

Method

Subjects.—The *S*s were 195 naive basic airmen trainees from Keesler Air Force Base

---

[1] This work was partially supported by a grant of the National Science Foundation to Edward A. Bilodeau, Principal Investigator.

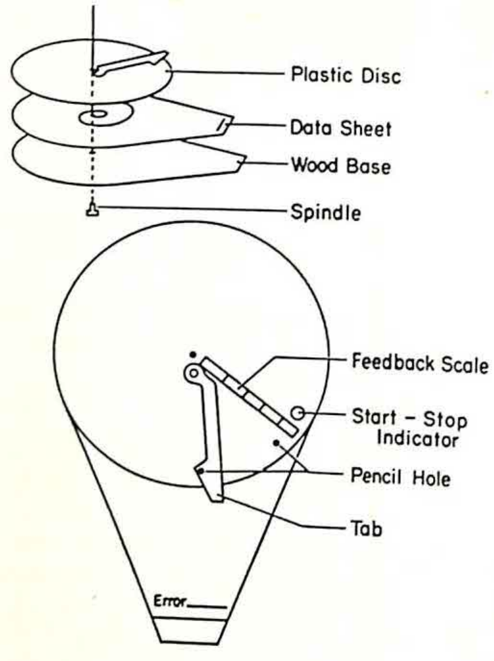


FIG. 1. A schematic representation of Arc Drawer A: A line-drawing device.

divided unsystematically into five groups.[2] Fifteen Ss were voided because of failure to follow instructions, and 19 Ss were excluded for failure to learn the task, leaving an average of 32 Ss per group.

Apparatus.—The apparatus was a dialing device developed by Levy (1961) for teaching S to draw a curved line of a given length. A schematic drawing is shown in Fig. 1. The essential parts include an opaque, plastic, disc through which S inserts his pencil, and from which he receives feedback after dialing. During the dialing, a tab covered the entire scale and withheld feedback until termination of the response. The feedback scale ran from −48, −47...0...+21, +22, and was devised to give both *magnitude* and *direction* of response error. Actually, the effect of dialing was to move a red line across the feedback scale. To assure that each S received KR, he was required to record his own error on the data sheet.

The drawing of a 16-in. line produced a terminal feedback error of zero. In order to accomplish this, S had to learn to turn through an arc of 255°. This 255° response, called R_a, was followed by KR, and constitutes an R_a-KR trial. Between R_a-KR trials all Ss were required to make an interpolated response, called R_b, by turning until a mark indicating *stop* appeared in a window. As a consequence, S made no error on R_b trials and no KR was administered.

Procedure.—All Ss were given 13 trials of R_a-KR alternating with 12 interpolated trials of R_b, making the sequence of events R_a-KR_a, R_b, R_a-KR_a, etc. The starting position was varied from trial to trial to prevent S from learning to go to a particular place instead of learning to dial through an arc of 255°.

Five groups were used and they differed only in the magnitude and direction of R_b. The groups are represented schematically in Fig. 2. Four groups turned in the counterclockwise direction through arcs of 255° or 235° or 205° or 165°. The fifth group responded with an arc of 255° in the clockwise direction and this turn was designated as an opposite response. For the four other groups turning in the counterclockwise direction, there was one group with R_b identical to R_a and three groups with R_b quantitatively similar to R_a.

The Ss were run in groups of 15 to 20 with an assistant E for each set of 5 Ss. The assistant Es aided in the changing of data sheets. The E explained the task and demonstrated the standard speed and direction of movement. Following the demonstration, the Ss were required to execute a practice response and any difficulties and misinterpretations were corrected. No S was instructed about the magnitude of R_b or R_a.

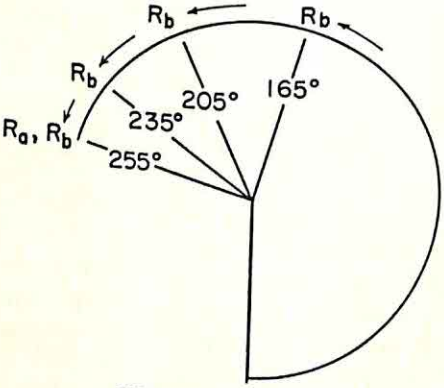


FIG. 2. A schematic representation of the experimental groups. (A final group, not shown, with R_b of 255° turned in the clockwise direction.)

---

[2] The cooperation of the Commander and of Marvin Karp of Keesler Air Force Base are gratefully acknowledged.

TABLE 1

MEAN ABSOLUTE ERROR AND VARIANCE OF RESPONSE IN DEGREES AS A FUNCTION OF FEEDBACK TRIALS

Groups		R_a-KR Trials							
		1	2	3	4	5	6	7	8
255° cc	Mean σ^2	155 637	65 476	39 251	36 353	30 158	25 83	22 52	23 176
235° cc	Mean σ^2	139 714	74 541	56 429	31 76	23 83	30 68	21 59	18 46
205° cc	Mean σ^2	134 854	70 634	41 425	31 223	38 116	28 123	25 83	31 157
165° cc	Mean σ^2	144 760	74 655	59 228	47 344	40 199	32 117	29 115	32 181
255° c	Mean σ^2	130 843	76 717	42 390	30 160	26 101	18 43	19 60	16 59

Note.—cc = counterclockwise; c = clockwise.

Results

Mean absolute errors and variances for the early R_a-KR trials are presented in Table 1. Both measures indicate that learning occurred rapidly. The only significant finding from a Groups × Trials analysis of variance was the Trials main effect. Also, an examination of algebraic error revealed no consistent difference between groups.

The results mean that various interpolated responses have no differential effect on the acquisition of the dialing response. The results do not show the *direction* of influence of the various interpolated R_b's. It is possible that all five interpolations produced the same amount of interference, or the same amount of facilitation, or no effect at all.

A second experiment was designed to determine the direction of influence of R_b. To this end, a comparison was made between a control group which made no R_b at all, and an experimental group which was required to make R_b.

EXPERIMENT II

Method

Subjects and apparatus.—A total of 60 men and 20 women Tulane undergraduates were unsystematically divided into two groups. Six *S*s were voided for failure to follow instructions, and 1 *S* was voided for failure to learn the task. There remained a total of 37 *S*s in the control group and 36 *S*s in the

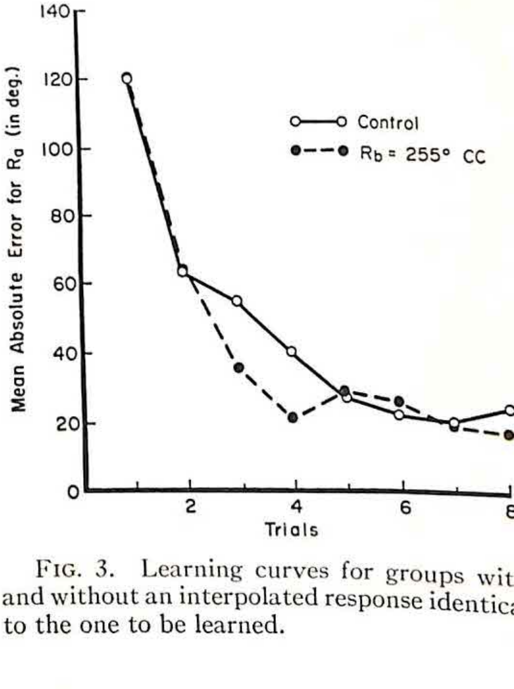


FIG. 3. Learning curves for groups wit and without an interpolated response identica to the one to be learned.

TABLE 2

VARIANCE OF RESPONSE IN DEGREES AS A FUNCTION OF FEEDBACK TRIALS

R_a-KR Trials	Group	
	Control	Experimental
1	899	678
2	473	563
3	308	125
4	269	94
5	96	136
6	75	85
7	74	74
8	82	44

experimental group. Approximately one-fourth of each group was composed of women.

Procedure.—The details are the same as for Exp. I except that the trials were reduced to 15 in number. For the experimental group, R_b was a guided counterclockwise response through an arc of 255°, a response identical to the R_a *S* was required to learn. The control group made no response instead of R_b. During the rest, the *S*s were required to sit quietly at their places.

Results

The results for R_a trials are shown in Fig. 3 and Table 2. Mean absolute error decreases rapidly for both groups with the greatest reduction in error occurring during the first few trials. Despite the discernible superiority of the experimental group between Trials 2 and 5, the difference between Trials 1 to 8 is not significant. Again, the main effect for Trials was the only significant finding. As in Exp. I the acquisition function was negatively accelerated and decreasing and the form of the curve was completely determined by the R_a-KR trials.

DISCUSSION

Experiments I and II show that the acquisition of the dialing response was not affected by the addition of R_b responses, whatever their length. Two other studies, where the interpolation of a response either before or after KR did not affect learning, were cited earlier. It had been concluded that absolute frequency of KR trials mattered, not relative frequency. In both experiments, however, the interpolated responses resembled the training responses and were in no way systematically controlled.

There is some evidence that interpolated responses are effective when subject to KR. In an experiment by I. McD. Bilodeau (1956), where a trials-delay technique was used, similar responses *and* KRs interpolated between R_a and KR_a did retard acquisition. She found a negative relationship between performance and increasing number of trials-delay. This could supply an explanation for the null data in the present experiment. Possibly, for an interpolated response to be effective, an R_a, KR_a, R_b, KR_b event sequence should be scrambled and/or displaced.

The present experiments bear on the response dimension of the Osgood Surface (1949). According to Osgood's empirical law for response variation: "where stimuli are functionally identical and responses are varied, negative transfer and retroactive interference are obtained, the magnitude of both decreasing as similarity between the responses increases." In Exp. I, where the groups differed only in respect to the similarity of the interpolated response to the goal response, there was no evidence for these effects. The empirical law also suggests that similar responses should aid learning, with the greatest amount of facilitation occurring when an identical response is used. The outcome of Exp. II does not support this prediction either.

A particularly interesting comparison with the Osgood Surface involves the group making an opposite interpolated response. Following the law for response variation, one would predict negative transfer, yet this group showed no evidence of negative transfer. A comment concerning the label "opposite" seems in order. One group was given this label because its R_b was of the same magnitude as the goal response, but in the opposite direction. We do not know if Osgood's designation implies oppositeness in all respects. Possibly there is no group which is opposite to another in all characteristics.

Detailed examination of the Opposite group shows that the so-called opposite response was in many respects similar to or the same as the goal response and one can wonder why positive transfer was not found. For example, the *S*s used the same posture and component acts on all trials. Many similar cues and skills are involved in making both the counterclockwise and the clockwise response of 255°. A similar conclusion was elaborated by Lewis, McAllister, and Bechtoldt (1953a, 1953b) in studies using a type of complex coordinator. They concluded from correlational analyses that the same skill was learned in original learning and interpolated learning. The conclusion held true even when a reversed task was used in interpolated learning. Data such as these have created considerable doubt concerning the appropriateness of the label "opposite." If the label were changed to "reversed direction," the naming would be more appropriate, but of course the finding would remain the same, i.e., no negative transfer.

Another possible interpretation of the null findings is that the amount of positive or negative transfer obtained depends on task difficulty. This would imply that, even with identical stimuli, the manipulation of the response dimension in conjunction with a simple skill would produce small, but unreliable, effects. Therefore, the failure to obtain reliable differences in the present experiments could be due to the simplicity of the task. As a technical matter, the present task is relatively easy and so is more favorable to a finding of negative transfer than to a finding of positive transfer.

In summary, this study casts additional doubt on the possibilities of producing retroactive interference in the learning of simple, discrete motor skills.

REFERENCES

ANNETT, J. Learning a pressure under conditions of immediate and delayed knowledge of results. *Quart. J. exp. Psychol.*, 1959, **11**, 3-15.

BILODEAU, E. A., & BILODEAU, I. McD. Variable frequency of knowledge of results and the learning of a simple skill. *J. exp. Psychol.*, 1958, **55**, 379-383.

BILODEAU, E. A., & BILODEAU, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243-280.

BILODEAU, E. A., BILODEAU, I. McD., & SCHUMSKY, D. A. Some effects of introducing and withdrawing knowledge of results early and late in practice. *J. exp. Psychol.*, 1959, **58**, 142-144.

BILODEAU, E. A., & RYAN, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, **59**, 414-419.

BILODEAU, I. McD. Accuracy of a simple positioning response with variation in the number of trials by which knowledge of results is delayed. *Amer. J. Psychol.*, 1956, **69**, 434-437.

LARRÉ, E. E. Interpolated activity before and after knowledge of results. Unpublished doctoral dissertation, Tulane University, 1961.

LEVY, C. M. Learning of a motor task with action and learning feedback. Unpublished master's thesis, Tulane University, 1961.

LEWIS, D., MCALLISTER, D. E., & BECHTOLDT, H. P. Correlational analysis of the learning and relearning of four different tasks on the Modified Mashburn Apparatus. *J. Psychol.*, 1953, **36**, 83-109. (a)

LEWIS, D., MCALLISTER, D. E., & BECHTOLDT, H. P. Correlational study of performance during successive phases of practice on the standard and reversed tasks on the SAM Complex Coordinator. *J. Psychol.*, 1953, **36**, 111-126. (b)

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132-143.

(Received July 20, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 520–521

SUPPLEMENTARY REPORTS

NET APPROACH GRADIENT IN APPROACH-AVOIDANCE CONFLICT

JAMES G. MARTIN, EVELINE M. LOEWE, ALLAN E. HINKLE, AND MARILYN L. FITZGERALD

Chico State College

This experiment tested the implication from conflict theory that net approach tendencies decrease nearer the goal in approach-avoidance. 16 albino rats, trained previously to run to food in an alley while 45 hr. food deprived, were shocked near the goal box. 2 days later Ss were tested for approach tendencies at points far (F) and near (N) the goal by measuring strength of pull. Each S pulled for 4 trials, half in FNNF order, half in NFFN order. Mean pull in grams at the near point was significantly less than at the far point ($p < .01$), as predicted.

A derivation from conflict theory not directly tested in published research is that, up to the point where the gradients intersect in approach-avoidance conflict, the net approach tendency decreases as the goal is neared (Miller, 1959). The present study was designed to test this derivation with the strength-of-pull technique used by Brown (1948) to verify some of the assumptions upon which the theory is based.

Method.—The *S*s were 20 male albino rats approximately 70 days old at the beginning of training. Sixteen *S*s survived the 3 mo. of training.

The apparatus was a metal alley approximately 160 cm. long, 15 cm. wide, and 15 cm. high, painted black inside with a hinged transparent plastic lid. Ten centimeters from the goal end of the alley was an electric grid 33 cm. long. The grid was raised $1\frac{1}{4}$ cm. from the floor.

A gram spring was mounted horizontally outside the start end of the alley. Attached to the spring was a piece of nylon fishline with two knotted loops placed to extend through a hole in the start end to a point 92 cm. (far point) or 46 cm. (near point) from the goal end of the alley. A rat in a cloth harness then could be attached to the far or near loop by means of a safety pin. A slider made of copper wire was mounted on the face of the spring to register the maximum excursion of the spring indicator for any given trial.

During 2 wk. of preliminary training *S*s were handled and permitted to explore the alley daily. During training to approach, *S*s were placed on an approximate 45-hr. food deprivation schedule. Each *S* ran six trials per training session to the food cup at the goal end of the alley, where he fed for 5 sec. on wet mash made from Purina laboratory pellets. This training continued every 2 days for 10 wk. During the last 2 approach-training days, the rats who had not reached roughly asymptotic running times were given extra trials. All *S*s were fed for an hour after each approach-training session.

Avoidance training, under the usual 45-hr. deprivation schedule, was accomplished during Training Session 36. Each rat ran six trials to the grid, activated by 12 v. from an inductorium, where he was shocked. Approximately 3 sec. after touching the grid he was removed from the alley. Some *S*s advanced toward the goal regardless of shock, but none were permitted to feed in the goal box. All *S*s approached and encountered the grid during each avoidance trial. The *S*s were fed for an hour after avoidance training.

Testing was accomplished 45 hr. later, during Session 37. After being placed in a cloth harness which was then attached to the nylon fishline, each *S* ran four successive approach trials, two to the near point and two to the far point. Half the *S*s, selected randomly, ran in the near-far-far-near order; the other half ran in the far-near-near-far order. The *S*s from each order were tested alternately. Once *S* ran far enough to activate the spring indicator he was allowed to pull for 5 sec. before being removed from the alley. Each pull was recorded to the nearest 10 gm. None of the strength-of-pull measurements was zero value; each *S* pulled against the spring during each test trial.

Results.—Mean pulls for the far-near-near-far group were 82.5, 62.5, 63.8, and 87.5 gm., respectively; for the near-far-far-near group, 65.0, 92.5, 91.3, and 67.5 gm. The two near and two far pulls were summed for each *S* and an analysis of variance was conducted. The main effects of counterbalancing condition and the interaction with distance were insignificant. Distance effects were significant, $F(1, 14) = 11.89$, $p < .01$. These observations lend support to the deduction that combined approach and avoidance tendencies result in a net decreasing approach gradient within the range of conditions tested.

REFERENCES

BROWN, J. S. Gradients of approach and avoidance responses and their relation to level of motivation. *J. comp. physiol. Psychol.*, 1948, **41**, 450–465.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959. Pp. 196–292.

(Received July 5, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 5, 521–522

STIMULUS-RESPONSE CODING AND AMOUNT OF INFORMATION AS DETERMINANTS OF REACTION TIME[1]

SIDNEY HELLYER

Defence Research Medical Laboratories, Toronto, Canada

Earlier investigators have shown reaction time and stimulus information to be linearly related with a slope dependent upon amount of pre-experimental practice. 5 housewives were run on 3 tasks varying in pre-experimental practice; light-naming, reading names from slides, and number-naming. Amount of information was varied from 1 to 3 bits. The results confirmed the linear relationship with the slope a decreasing function of pre-experimental practice. However, even after 5040 trials on the task with the least such practice, light-naming, differences in reaction time corresponding to differences in information amount had decreased but little, indicating that the information effect is stable, at least for relatively unusual tasks.

Hyman (1953) demonstrated that, for complexly coded stimulus-response combinations, reaction times show a linear increase with an increase in the average amount of information in the stimulus. Mowbray (1960), using a simple number-naming task, found reaction time to be constant, and not dependent upon stimulus uncertainty. He suggested that the apparent disagreement could be accounted for on the basis of prior practice.

The present experiment reproduced the conditions of both the preceding experiments with a single group of *S*s. Furthermore, it included a third task which would seem to be between those of Hyman and Mowbray with respect to amount of prior learning. Also, learning effects on Hyman's more complex task were studied directly.

Method.—The *S*s were 5 paid housewives. Three displays were used: (*a*) Hyman's (1953) panel of eight lights, the names of the combinations being those used by Hyman, (*b*) a Burroughs "Nixie" numerical read-out tube as used by Mowbray (1960), and (*c*) a display using eight slides each of which showed one of Hyman's names.

First, the number and slide displays were used with three amounts of information, i.e., two, four, or eight equiprobable stimuli per set, or 1, 2, or 3 bits of information per stimulus presentation for a total of 6 sessions. Second, the more complex light-naming task was used with the same three amounts of information randomized over blocks of 3 sessions for a total of 21 sessions. Finally, the slide and the light displays were used with the amount of information being varied by altering both the number of possible stimuli and their conditional probability of occurrence. Four such relations were selected from those used by Hyman (1953), i.e., those for average information values of 0.72, 1.36, 1.79, and 2.21 bits. This involved 8 sessions.

To each *S* a single condition was presented on any one day and the order of presentation of stimuli was randomized for each daily session. Reaction times were measured with a Hunter Model 120A Klockounter which started when the stimulus appeared and was stopped by a voice key when *S* spoke. There were 240 reaction times measured during each 1-hr. session—30 with each of the eight numbers or words.

---

[1] Defence Research Medical Laboratories Project No. 246, DRML Report No. 246-13, PCC No. D77-94-20-46, H. R. No. 236.

The experimental procedure differed from Mowbray's in two ways. All 5 *S*s were run with every amount of information. Also, the *S*s were presented all numbers from 1 to 8 inclusive for all information conditions. For example, in the two-choice situations, a 1-hr. session consisted of four blocks of 60 trials each, with a different randomly-selected set of numbers for each block.

Results.—The group mean reaction times for each of the three display conditions are plotted in Fig. 1 as a function of amount of information per stimulus presentation. Amount of transmitted information is almost identical with the amount of information per stimulus presentation since the *S*s had a very low error rate: less than 1% for all *S*s and conditions with the exception of 1 *S* with an error rate of slightly more than 1.5% for the difficult light-naming task.

The results for the light-naming task agree very well with those obtained by Hyman (1953). All but one of the sets of data of individual *S*s show a linear relation between reaction time and amount of information. For this 1 *S*, the trend was very similar to the others but was significantly nonlinear.

The results for number-naming were similar to Mowbray's (1960) in that the relation was linear and the effect of increasing the amount of information was very slight. Unlike Mowbray's results, there was a significant increase in reaction time corresponding to an increase in amount of information.

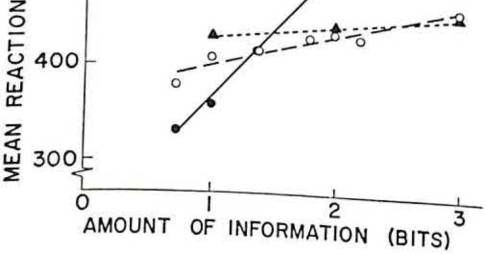


FIG. 1. Group mean reaction time as a function of the amount of information in the stimulus for the three displays: Light, Slides, and Numbers.

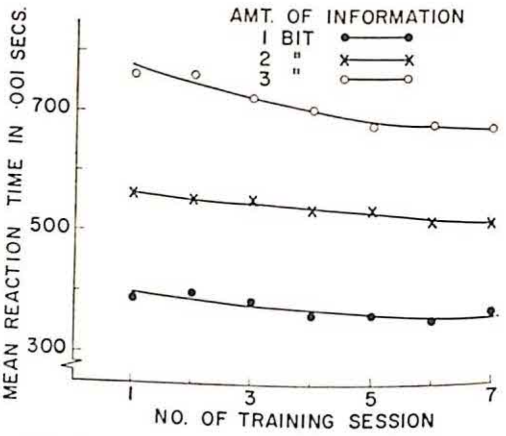


FIG. 2. Group mean reaction time as a function of the amount of information in the stimulus and the number of training sessions.

As predicted, the group data for the slide display appear as a straight line lying between those for the other tasks. However, the slope of this line is not significantly greater than that for the number-naming task.

Figure 2 shows the results for the seven blocks of learning trials with the light-naming task. An analysis of variance showed that a significant amount of learning occurred and that there was more pronounced learning under conditions where more information was involved. However, even after 5,040 trials per *S*, differences in reaction time corresponding to differences in amount of information had decreased but little from their initial values, indicating that the amount of information in the stimulus has a stable and important effect on reaction time, at least for relatively unusual tasks.

REFERENCES

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.
MOWBRAY, G. H. Choice reaction times for skilled responses. *Quart J. exp. Psychol.*, 1960, **12**, 193–202.

(Received June 25, 1962)

NEW

DIGITAL TEST MODULES
Announced by
Scientific Prototype

Scientific Prototype announces new Digital Test Equipment Building Blocks which provide researchers with a modular system with which they can quickly and inexpensively assemble equipment for random and fixed ratio programs, counting, measuring, decision making, etc. This new research tool brings new speed, economy, and flexibility to your laboratory.

Scientific Prototype now offers a revolutionary new series of Digital Test Modules for use in the measurement and control of experimental variables. A small number of easy-to-use building blocks will perform any task. No special training is required.

FEATURES: The new units are entirely transistorized. They operate in decimal arithmetic, and provide direct numerical indication. All modules have buffered outputs which provide high fan-out ratios and permit outputs to be interconnected for added flexibility. In addition, any input or output may be short-circuited without damage to the equipment.

An internal 10 KC clock with an accuracy of 0.001% permits the precise measurement of reaction times, response rates, etc.

A complete line of input and output accessories including digital printers, analog-to-Digital converters, high power load switching equipment, and tape readers and punches is available.

Write now for full information and to reserve your copy of "Digital Techniques", our notes and comments on equipment and techniques.

SCIENTIFIC PROTOTYPE MFG. CORP.

Behavior Analysis Apparatus . . . operant conditioning kits, rodent test cages, pellet dispensers, soundproof chambers, recorders, noise generators, shock sources, times, switches, digital test equipment, tachistoscopes and special systems.

623 W. 129th STREET, NEW YORK 27, NEW YORK
TELEPHONES (212) MO 6-4674 and 4789

the monthly review journal
of the
American Psychological Association

CONTEMPORARY
PSYCHOLOGY

reviews

criticism

opinion

on

the latest books, films,

and other instructional media

lists of new books

letters-to-the-editor

$1.00 a copy

$10.00 a year
($10.50 foreign)

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333-16th Street, N.W.
Washington 6, D.C.

Educational Psychology in the U.S.S.R.

EDITED AND WITH AN INTRODUCTION BY
BRIAN AND JOAN SIMON

Prepared with the active cooperation of the Institute of Psychology of the Academy of Educational Sciences in Moscow, this volume provides a unique survey of recent Soviet work on the psychology of learning. All of the papers are by leading Soviet psychologists. $6.75

Mathematical Methods in Small Group Processes

EDITED BY JOAN CRISWELL, HERBERT SOLOMON,
AND PATRICK SUPPES

These 22 papers, presented at a symposium held at Stanford University in June 1961, represent a varied sampling of mathematical methods applied to problems of interpersonal behavior in small groups. The material includes new applications of statistical learning theory, of the Lorge-Solomon probability model of group problem solving, and of signal-detection theory. New expressions are given for power relationships, subtask phasing, non-oriented behavior, and a new topological measure of "closeness" applicable to a variety of data. *Stanford Mathematical Studies in the Social Sciences, VIII.* $9.75

Social Pressures in Informal Groups

LEON FESTINGER, STANLEY SCHACHTER, AND KURT BACK

This book, first published in 1950, has become generally recognized as one of the important pioneering works in the field of social psychology. "For theory of group structure and functioning, a number of very suggestive points emerge. . . . The study shows an impressive use of a battery of techniques in a carefully planned and systematic way."—*Human Relations* $5.00

Order from your bookstore, please

STANFORD UNIVERSITY PRESS

WILEY BOOKS

JOHN WILEY & SONS announces
the publication of a major research and reference work . . .

HANDBOOK OF MATHEMATICAL PSYCHOLOGY

In three volumes

Edited by R. DUNCAN LUCE, University of Pennsylvania, ROBERT R. BUSH, University of Pennsylvania, and EUGENE GALANTER, University of Washington

Although mathematics has been used by psychologists for many years, the intensive development of mathematical theories of behavior is relatively recent. The field of mathematical psychology has progressed to such an extent that a knowledge of it is a "must" for many students and professional psychologists. The three editors of the HANDBOOK OF MATHEMATICAL PSYCHOLOGY recognized this need and have prepared the first systematic survey of the major mathematical applications to psychology.

VOLUME I of the HANDBOOK includes measurement, psychophysics, reaction time, computers, and statistics; VOLUME II presents mathematical learning models, formal theories of language, and models for social interaction; and VOLUME III is planned to contain chapters on visual and auditory mechanisms, utility and preference theory, discrimination learning, and mathematics. The three volumes of this work offer the expert a compact and convenient summary of research as well as an extensive bibliography, and provide the student with a high-level introduction to the field.

VOLUME I is now available—it contains 491 pages and is priced at $10.50.

VOLUME II will be available soon—it will contain approximately 624 pages and will probably be priced at $12.00.

VOLUME III is in preparation.

the ideal reference volume and companion to the HANDBOOK . . .

READINGS IN MATHEMATICAL PSYCHOLOGY

In two volumes

Edited by R. DUNCAN LUCE, ROBERT R. BUSH, and EUGENE GALANTER

Serving the dual purpose of a companion to the HANDBOOK and a valuable reference source, READINGS IN MATHEMATICAL PSYCHOLOGY contains a wide variety of research papers reprinted from professional journals. Each paper represents a major contribution to the progress of mathematical psychology in the last fifteen years.

VOLUME I contains papers on measurement, psychophysics, reaction time, learning, and stochastic processes; VOLUME II is planned to contain papers on utility and preference theory, language, computers, social interaction, and auditory and visual mechanisms.

VOLUME I will be available soon—it will contain approximately 580 pages and will probably be priced at $12.00.

VOLUME II is in preparation.

two books of related interest . . .

CONCEPT LEARNING: An Information Processing Problem

By EARL B. HUNT, University of California. 1962. 286 pages. $7.50.

EXPLORATIONS IN COGNITIVE DISSONANCE

By JACK W. BREHM, Duke University, and ARTHUR R. COHEN, New York University. 1962. 334 pages. $7.95.

note change of address

JOHN WILEY & SONS, Inc.
605 Third Avenue
New York 15, N.Y.

Journal of
Experimental Psychology

Vol. 65, No. 6 June 1963

GUSTATORY ADAPTATION TO SALIVA AND SODIUM CHLORIDE[1]

DONALD H. McBURNEY and CARL PFAFFMANN
Brown University

Human NaCl thresholds were measured under 2 conditions: (a) salivary influence excluded by rinse with distilled water or 1 of 3 weak concentrations of NaCl between stimulations, and (b) salivary influence maximized by using no rinse. Adaptation to distilled water yielded a median threshold for 4 Ss of .00014 M vs. .0043 M for adaptation to saliva. The latter value approximated the resting concentration of sodium in the saliva of these Ss. Adaptation to saline solutions raised the thresholds to values slightly above each of the adapting concentrations. It is concluded that salt thresholds normally reflect the state of adaptation to salivary sodium. Changes in general physiological state of the organism which affect salivary composition might be expected to shift the taste threshold by a process of sensory adaptation.

When human *S*s are depleted of sodium chloride, the taste threshold for this compound is lowered (de Wardener & Herxheimer, 1957; Henkin & Solomon, 1962; Yensen, 1959), but the mechanism responsible for the change has not been clarified. Earlier writers (Hartridge, 1945; Richter, 1942; Richter, Holt, & Barelare, 1938; Young, 1938, 1941, 1949) proposed that lowered sodium content of the blood, after adrenalectomy, for example, might affect the sensitivity of the salt receptors directly to cause a lowered threshold. Behavioral determinations of salt thresholds in rats before and after adrenalectomy (Carr, 1952; Harriman & MacLeod, 1953) and electrophysiological recording of thresholds in normal and adrenalectomized rats (Pfaffmann & Bare, 1950), however, showed a change only in preference threshold, not in absolute sensitivity. Thus, it is commonly concluded that a central, rather than a receptor, process is responsible for these threshold effects (Pfaffmann, 1957).

That the sensitivity of the receptor may be a function of the sodium chloride level of the fluids bathing the taste bud has often been suggested. Liljestrand and Zotterman (1954, p. 292) noted that according to the older literature, the flat taste of water was held to be a distinctive *sapor*

---

[1] This study was supported in part by grants-in-aid from the General Foods Corporation and the National Science Foundation.

The authors wish to express their appreciation to Trygg Engen and John Kraukopf for suggestions and advice in psychophysical procedures and data analysis.

insipidus or *sapor aquosus* characteristic of solutions containing less salt than the saliva. Richter and MacLean (1939) remarked on the possibility that the taste threshold might be determined in part by the salt content of the saliva (p. 6). Zotterman (1958) also observed:

> After a water rinse even as weak as a solution as 0.003 M NaCl caused an obvious and persisting response in [the rat chorda tympani]. . . . The chloride concentration of the rat's saliva seems to be a constant stimulus to the salt receptors (p. 523).

Adapting the tongue to relatively high concentrations of salt elevates the human threshold to a value slightly above the concentration of the adapting solution (Abrahams, Krakauer, & Dallenbach, 1937; Hahn, 1934). It is also known that salivary sodium level drops under salt deprivation (McCance, 1938; Thorn, Schwartz, & Thaysen, 1956). These considerations suggest that the salt taste threshold may be a function of the sodium content of saliva to which the taste buds are normally adapted under physiological conditions. The present study was designed to obtain evidence relative to this hypothesis.

METHOD

Subjects and apparatus.—Five male undergraduates were paid to serve as *Ss*. All solutions used in the experiment were made with distilled water and were maintained by a water bath at 34° C., a temperature which yielded no temperature sensation when the solutions were applied to the *S*'s tongue. The *S* sat with the tongue extended slightly

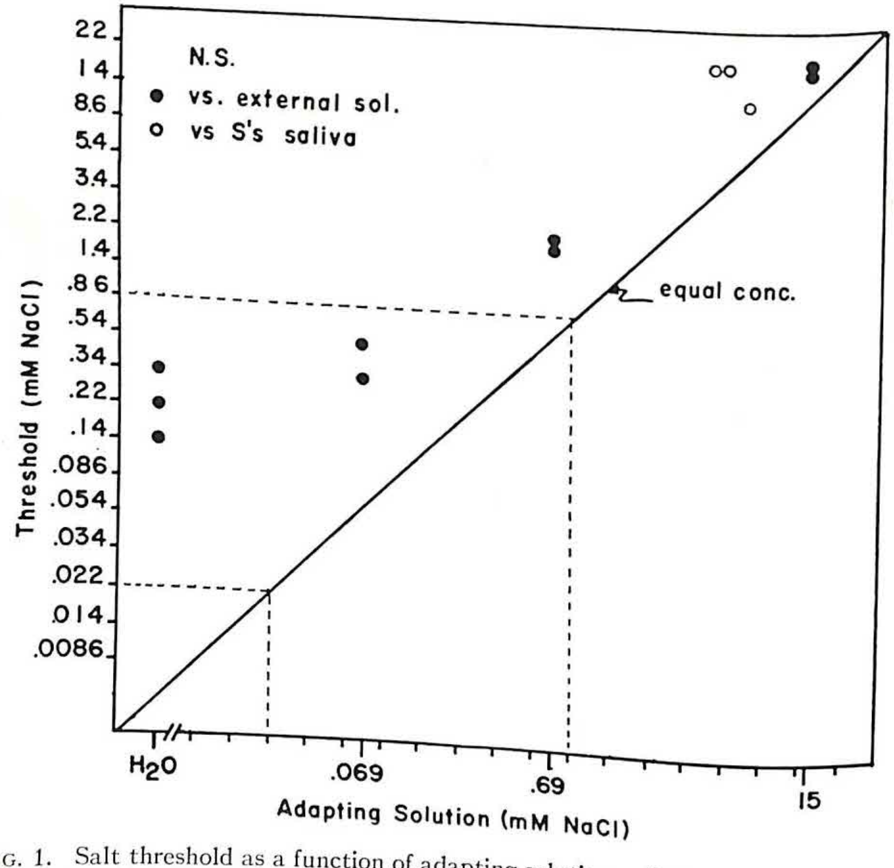


FIG. 1. Salt threshold as a function of adapting solution. (Subject NS: ordinate, threshold; abscissa, adapting solution or salivary sodium concentration in mM.)

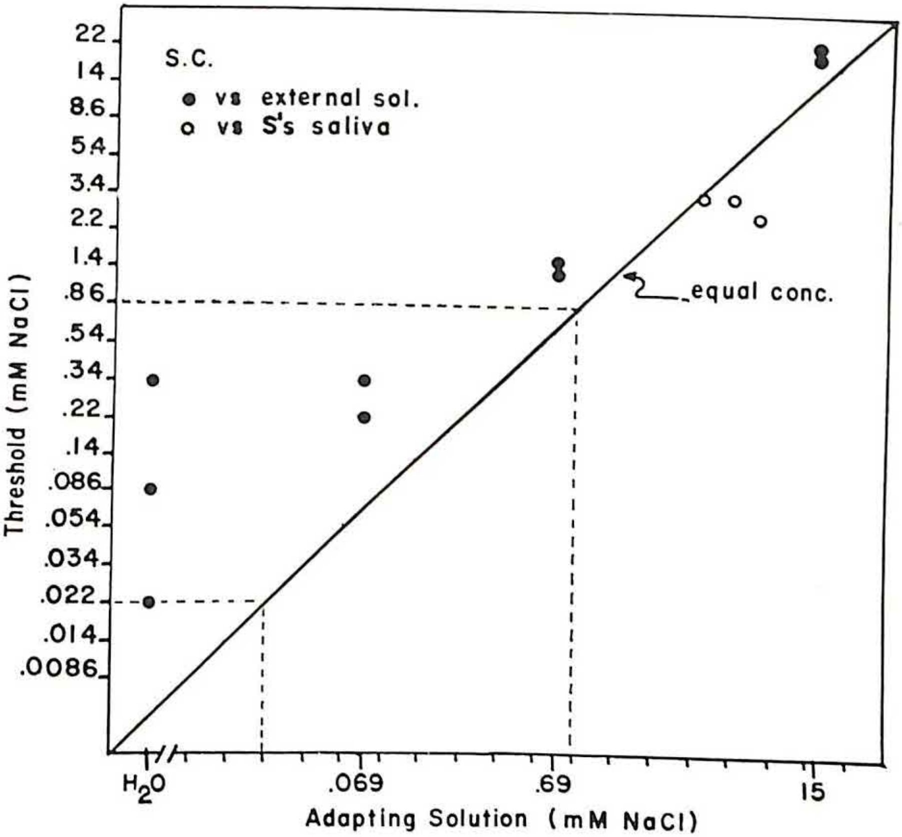


FIG. 2. Subject SC, same as Fig. 1.

between the lips so that a gravity flow system could deliver a continuous stream of solution over the anterior dorsal tongue surface and then into a collecting funnel beneath the chin. The tongue, therefore, was constantly bathed by an externally delivered solution and entirely free from any contact with the saliva except under one condition where no adapting rinse was employed. In this case, the tongue remained in the mouth between stimulations and, therefore, was adapted to saliva. Both the adapting and the stimulus solutions irrigated the entire exposed surface. The *S*s were able to remain comfortably in this situation for an extended period of time.

Procedure.—Threshold determinations were made as follows. The rinse flowed over the tongue for at least 2 min. before the first stimulus was presented. The *S* moved his tongue briefly out of the adapting stream or extended it into position for the stimulus, replacing it after each judgment. The stimulus consisted of 3–5 ml. of salt solution presented by pipette. An interstimulus interval of at least 15 sec. was employed with a modified ascending method of limits. This method, which was a type of tracking procedure, was chosen in order to follow any changes in threshold which might occur during sessions. Determinations were continued for 25 min. of adaptation and stimulation under each condition, yielding approximately 10 values. The first stimulus identified as salt was considered the threshold for a particular determination. Distilled water was presented as a "blank" about every third stimulus, depending on *S*'s false alarm rate, i.e., positive calls given to water. No corrections were given, except for false alarms. Time of day was constant and *S*s refrained from eating or smoking 1 hr. before the session. Four *S*s were trained for 16 days before the start of the experiment; the fifth received 15 days training. Five adapting conditions were employed, covering the range from lowest threshold possible to somewhat higher than the normal threshold: (*a*) distilled water, (*b*) .000069 M NaCl, (*c*) .00069 M NaCl, (*d*) .015 M NaCl, and (*e*) in the mouth between stimuli (hereafter called "in-the-mouth" condition). Conditions 1 and 5 were presented three times and Cond. 2, 3, and 4, twice, during the 6 days of the experiment in a partially counterbalanced order, two conditions being used on each of the 6 days.

RESULTS

Results are summarized for each *S* in Fig. 1–4, which show threshold obtained under the different adapting conditions. Each point indicates the median determination for one condition for 1 day. The scales on both axes are the same, the diagonal describing equal concentrations of adapting solution and threshold solution. Thus, if the *S* were to respond to the first stimulus that was of a higher concentration than the adapting solution, for instance, the threshold would always be slightly above the diagonal. Open circles, representing in-the-mouth condition, are plotted at the equivalent sodium concentration of *S*'s saliva. Approximately 3–5 ml. of resting saliva were obtained from each *S* during the experimental session. Salivary sodium concentrations were determined by flame photometry. Individual false alarm rates were .07 for SC, .08 for NS, .11 for BB, .12 for RS, and .38 for HC. Because of the high rate of the last *S*, and the consequent unreliability of his threshold, his data will not be considered.

DISCUSSION

It is seen from the figures that the threshold is a function of the sodium chloride content of the adapting solution. The grand median for all *S*s with the water rinse is one quarter of a stimulus step above .00014 M (range .000022 M

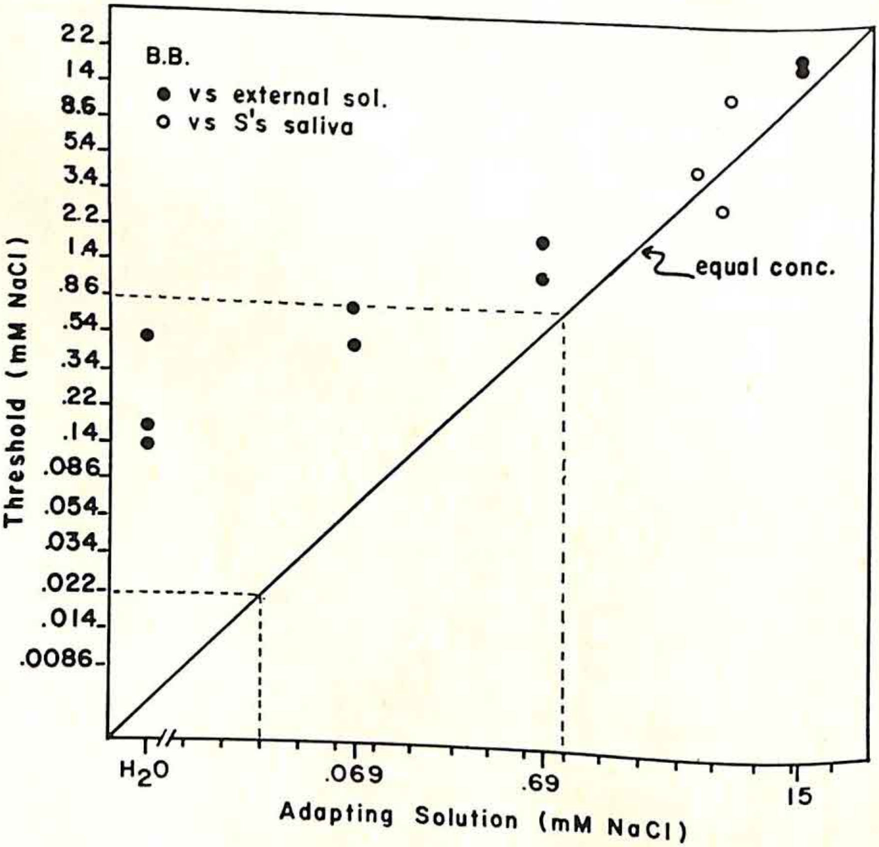


FIG. 3. Subject BB, same as Fig. 1.

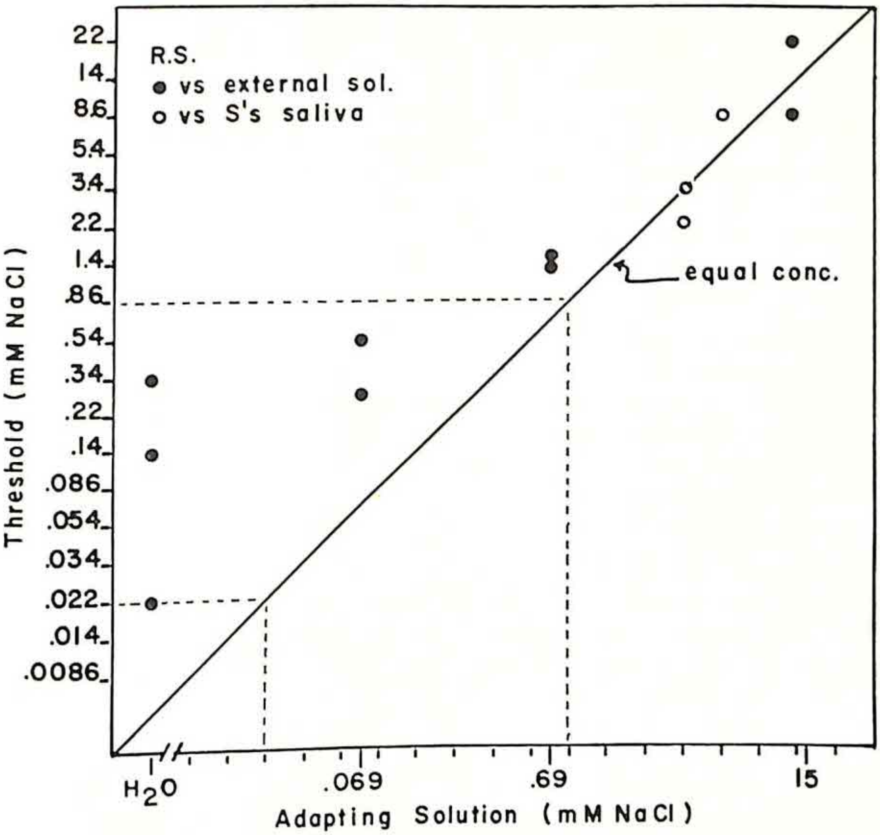


FIG. 4. Subject RS, same as Fig. 1.

to .00054 M). The threshold increases with increasing salt content of the rinse, .000069 M being sufficient to elevate the threshold above the value obtained with water. It will be noticed that the threshold is always above the level of the adapting solution (in the rinse condition) with the single exception of one daily threshold for 1 *S*. The minimum threshold of .00014 M compares with Yensen's (1959) minimum of approximately .00089 M obtained under salt deprivation, with .00086 M found by de Wardener and Herxheimer (1957) under conditions of greatly increased water intake, and with .0004 M found by Henkin and Solomon (1962) in patients having either untreated pituitary insufficiency or untreated, or inadequately treated, adrenal insufficiency. Thus, it appears that changing the adaptation level of the tongue in normal *S*s produces a threshold change of the same order of magnitude as the range of thresholds found in human *S*s depleted of sodium chloride.

The thresholds obtained under the "in-the-mouth" condition fell on the function relating thresholds to adaptation but with more variability and individual differences. It appears that the threshold under normal conditions is approximately that which would be expected if the sodium concentration were the major determinant of threshold. Sodium is, of course, only one ionic component of saliva besides potassium, calcium, chloride, phosphate, bicarbonates, etc. Hahn (1934) has shown that adaptation to potassium chloride did not affect the threshold for sodium chloride in man. Beidler (1953), in the rat, also found no cross adaptation between NaCl rinse and KCl or other monovalent salt stimuli. The threshold for sodium chloride solutions may, therefore, be expected to be relatively independent of the ambient level of other salivary ions. Other variables, also, may have been operative. Salivary sodium may have been fluctuating during the experiment, saliva being sampled once per session, or the flow of saliva may have been insufficient to adapt the tongue rapidly enough to maintain adaptation level between stimuli, since *S*s occasionally reported dryness of the mouth during the sessions.

Our salivary sodium concentrations are in agreement with those reported for resting mixed saliva (Schneyer & Turner, 1954). Since sodium concentration increases with rate of flow (Fenton, 1960), by taking small quantities of relatively unstimulated saliva we sampled the saliva which actually bathed the tongue in our experimental situation, yielding a concentration which, when plotted against threshold, fell on the function of threshold vs. external adapting solution.

Most taste threshold studies reported in the literature include a brief water rinse before each stimulus to free the mouth of saliva. In preliminary work with the same *S*s, we studied the effect of no rinse vs. a brief prerinse under the "in-the-mouth" condition. The no-rinse condition yielded a grand median one-quarter step below .0086 M, whereas prerinse gave a value of .0022 M. It would appear, then, that a brief rinse procedure would lower the threshold to some extent but would also introduce several uncontrolled variables, including volume of rinse, time held in the mouth, amount of mixing with saliva, and variable rate of salivary flow, all of which would affect the adaptation level and, hence, threshold in an unknown and variable manner.

In Pfaffmann and Bare's (1950) electrophysiological study on the normal and adrenalectomized rat, the tongue was rinsed between stimuli in tap water. This had the effect of eliminating salivary influence and adapted the threshold in both groups of animals in the experiment to the same low value. The results of the present experiment are consistent with their conclusion that the sensitivity of the receptor per se does not change under salt depletion but it would appear that the adaptation level of the receptor and, hence, threshold might change as a consequence of a change of the salivary sodium level.

Henkin and Solomon's (1962) patients with adrenal or pituitary insufficiency yielded lower salt thresholds than normals. Frawley and Thorn (1951) observed an inverse relation between adrenal corticoid level and salivary sodium concentration, which suggests that changes in salivary sodium and chloride are not responsible for the drop in threshold. That is, when cortical hormone is deficient, then secretion of NaCl is excessive in both kidney and salivary glands and a high level of salivary sodium might be expected. But this is true only if serum sodium is high. When serum sodium falls, then adrenal insufficiency would be associated with low salivary sodium.

Several ways of obtaining low thresholds may now be predicted with low salivary sodium levels: (*a*) low serum sodium and high corticoids (by salt deprivation), (*b*) very low serum sodium and low corticoids (by adrenal insufficiency or adrenalectomy with low salt intake), and (*c*) normal serum sodium and high corticoids. Normal salivary sodium and normal threshold may be predicted from: (*a*) normal serum sodium and corticoid level (in the normal situation) and (*b*) moderately low serum sodium and low corticoids (with adrenal insufficiency and adequate dietary salt to maintain only moderately low serum sodium level).

CONCLUSIONS

1. Continuous adaptation of the tongue to distilled water between stimulations yields very low taste thresholds for sodium chloride.

2. Adaptation of the tongue to low concentrations of sodium chloride elevates the threshold from its minimum to a value just greater than that of the adapting concentration of sodium chloride.

3. Plotting the normal threshold against sodium concentration of resting saliva shows

that the normal threshold can be included on the function relating adapting concentration to threshold.

4. The stimulus values encompassed by this effect covers the range of thresholds found in humans in a state of sodium chloride depletion or adrenal insufficiency. They are correspondingly of the same magnitude as reported in the adrenalectomized rat.

REFERENCES

Abrahams, H., Krakauer, D., & Dallenbach, K. M. Gustatory adaptation to salt. *Amer. J. Psychol.*, 1937, **49**, 462–469.

Beidler, L. M. Properties of chemoreceptors of tongue of rat. *J. Neurophysiol.*, 1953, **16**, 595–607.

Carr, W. J. The effect of adrenalectomy upon the NaCl taste threshold in rat. *J. comp. physiol. Psychol.*, 1952, **45**, 377–380.

de Wardener, H. E., & Herxheimer, A. The effect of a high water intake on salt consumption, taste thresholds and salivary secretion in man. *J. Physiol.*, 1957, **139**, 53–63.

Fenton, P. F. The mouth and esophagus. In T. C. Ruch & J. F. Fulton (Eds.), *Medical physiology and biophysics*. Philadelphia: Saunders, 1960. Pp. 917–923.

Frawley, T. F., & Thorn, G. W. The relation of the salivary sodium-potassium ratio to adrenal cortical activity. In J. R. Mote (Ed.), *Proceedings, second clinical ACTH conference: I. Research.* New York: Blakiston, 1951. Pp. 115–122.

Hahn, H. Die Adaptation des Geschmacksinnes. *Z. Sinnesphysiol.*, 1934, **65**, 105–145.

Harriman, A. E., & MacLeod, R. B. Discriminative thresholds of salt for normal and adrenalectomized rats. *Amer. J. Psychol.*, 1953, **66**, 465–471.

Hartridge, J. The importance of taste and smell in nutrition. *J. Physiol.*, 1945, **103**, 34–35.

Henkin, R. I., & Solomon, D. H. Salt taste threshold in adrenal insufficiency in man. *J. clin. Endocrin. Metab.*, 1962, **22**, 856–858.

Liljestrand, G., & Zotterman, Y. The water taste in mammals. *Acta physiol. Scand., Stockholm*, 1954, **32**, 291–303.

McCance, R. A. The effect of salt deficiency in man on the volume of the extracellular fluids, and on the composition of sweat, saliva, gastric juice and cerebrospinal fluid. *J. Physiol.*, 1938, **92**, 208–218.

Pfaffmann, C. Taste mechanisms in preference behavior. *Amer. J. clin. Nutr.*, 1957, **5**, 142–147.

Pfaffmann, C., & Bare, J. K. Gustatory nerve discharges in normal and adrenalectomized rats. *J. comp. physiol. Psychol.*, 1950, **43**, 320–324.

Richter, C. P. Self regulatory functions. *Harvey Lectures*, 1942, **38**, 63–103.

Richter, C. P., Holt, L. E., & Barelare, B. Nutritional requirements for normal growth and reproduction in rats studied by the self selection method. *Amer. J. Physiol.*, 1938, **122**, 734–744.

Richter, C. P., & MacLean, A. Salt taste thresholds of humans. *Amer. J. Physiol.*, 1939, **126**, 1–6.

Schneyer, L. H., & Turner, F. The sodium and potassium concentration of "resting" saliva from individual gland pairs in man. *J. dent. Res.*, 1954, **33**, 716.

Thorn, H. A., Schwartz, T. L., & Thaysen, J. H. Effect of sodium restriction on secretion of sodium and potassium in human parotid saliva. *J. appl. Physiol.*, 1956, **9**, 477–480.

Yensen, R. Some factors influencing taste sensitivity in man: II. Depletion of body salt. *Quart. J. exp. Psychol.*, 1959, **11**, 230–238.

Young, P. T. An hypothesis concerning the mechanism of appetite. *Psychol. Bull.*, 1938, **35**, 716–717.

Young, P. T. The experimental analysis of appetite. *Psychol. Bull.*, 1941, **38**, 129–164.

Young, P. T. Studies of food preference, appetite and dietary habit: V. Techniques for testing food preference and the significance of results obtained with different methods. *Comp. Psychol. Monogr.*, 1949, **19**, 1–58.

Zotterman, Y. Studies in the nervous mechanism of taste. *Exp. Cell Res.*, 1958, Suppl. No. 5, 520–526.

(Received July 30, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 530–536

ON THE PROBLEM OF HYSTERESIS IN PSYCHOPHYSICS[1]

HANNES EISLER AND CHRIS OTTANDER
University of Stockholm, Sweden *University of Uppsala, Sweden*

Subjective velocity was estimated by modified methods of magnitude estimation and category rating, presenting the stimuli in ascending and descending series separately. Hysteresis was found and the midpoints were predicted with good precision from a model describing the relation between category and magnitude scales. It was shown that the main factor producing hysteresis is a difference in the "subjective zero," ϕ_0, of the power function $\psi = k(\phi - \phi_0)^n$, between ascending and descending series. It is tentatively suggested that the same effect may account for the time-order error.

Hysteresis is an effect found most clearly in psychophysical bisection or equipartition experiments. If a certain range of a continuum is partitioned in equal intervals by an *O*, the position of the divisions obtained will depend upon whether the stimuli are presented in ascending or in descending order. Thus, in bisection the midpoint as set by *O* will be higher if the stimuli are presented in the order: lower fixed stimulus—adjustable stimulus—higher fixed stimulus, than if this order is reversed. Hysteresis has been demonstrated for subjective brightness, loudness, and weight (all so-called prothetic continua) and with some doubt for pitch. It is not found in typical metathetic continua. The phenomenon is described in more detail by Stevens (1957, 1961). Though disconcerting from the point of view of scaling theory, the effect of hysteresis has not yet been explained.

Recent findings by Eisler (1962a, 1962b, 1963a, 1963b) on the relation between magnitude and category scales provide a model which may be used to tackle the problem of hysteresis. According to this model, the category rating scale *K* and the magnitude estimation scale ψ observe the following relation

$$K = \alpha \log (\psi - \psi_0) + \beta \qquad [1]$$

The parameter ψ_0 constitutes the ψ intercept $-q/k$ of the straight line

$$\sigma = k\psi + q, \qquad [2]$$

the generalized form of Weber's law for subjective continua. Here σ denotes the intraindividual standard deviation of ψ.

The aim of the present study was: (*a*) to investigate whether the model mentioned above may be applied to a bisection task or whether it is suitable only in typical category estimation, and (*b*) to find out whether the hysteresis effect can be predicted within the frame of this model, e.g., by different ψ intercepts $-q/k$ for ascending and descending series.

These problems were studied on the continuum of subjective velocity as displayed by a luminous spot moving in a circular track. To summarize the results, Equations 1 and 2 permitted good prediction of the midpoints. The hysteresis effect, however, was due not to a difference in the ψ intercepts but to a difference between the ascending and descending magnitude scales.

---

[1] This research has been supported by Aerospace Research, United States Air Force, through its European Office.

METHOD

Apparatus

The apparatus was built to display a luminous spot (diameter .5 cm.) moving in a circular track (diameter 24 cm.). The luminance of the spot was 46.67 ft-L and it appeared on a screen with a luminance of .0060 ft-L (luminance measured with a Spectra brightness spotmeter Model UB).

A synchronous motor with a continuous gear made it possible to vary the angular velocity between 10 and 220 rpm. The gear consisted of two conical shafts side by side in reversed positions one driving the other via a wheel between the shafts. The velocity was varied by sliding the wheel along the shafts. The velocity was read off a calibrated meter coupled to a miniature motor run as a generator by the output shaft. The luminous spot was provided by a lamp attached to the end of an arm fixed at right angles to the output shaft. As the motor was run at a constant speed, *O*s got no secondary auditory clues as to the velocity of the spot. The presentation time and the interval between presentations were kept constant by a timer system.

The *O* was seated 2.5 m. from a semitransparent screen (87.5 × 87.5 cm. inclusive wooden frame of 3.5 cm.) showing the luminous spot. There were no landmarks on the screen. The experimental room was kept in semidarkness during sessions.

Procedure

Two experiments were carried out, using modifications of the methods of magnitude estimation and category rating. Magnitude estimation preceded category rating by at least 1 day and the same 12 *O*s took part in both experiments. The same 15 stimuli (see Table 1) were used in both.

Experiment I: Magnitude estimation.—In this investigation it was essential to present the stimuli in ascending and descending series separately. If the same sequence were shown repeatedly, *O*s would probably start learning number responses. In order to avoid this source of error, combinations of 6, 7, or 8 stimuli were drawn at random from the total of 15 and presented in ascending or descending order. Eight ascending and eight descending series were carried out so that, all in all, eight estimates were obtained for each stimulus from each *O*—four belonging to an ascending and four to a descending series. The eighth stimulus (82 rpm) served as standard and was assigned the value 10. The standard was shown once (for 60 sec.) just before the beginning of the first magnitude estimation series.

In the magnitude estimation series each stimulus was presented for 4 sec. with an interval of 5 sec. between each presentation.

Experiment II: Category rating.—In order to make the bisection procedure as comparable to magnitude estimation as possible, an estimation rather than a production method was chosen. (With magnitude *production* it would have been impossible to compute the standard deviations needed for the test of the model.) On the other hand the procedure was made as similar as possible to the typical bisection experiment. At the outset, each *O* was shown the lowest and highest stimuli (20 and 200 rpm) for 30 sec. each and informed that the numbers 1 and 3 corresponded to these. The *O* was then shown the lowest stimulus again for 4 sec. and after an interval of 5 sec., *O* was shown one of the remaining 13 stimuli drawn at random which he had to assign to one of the subjectively equally spaced categories 1, 2, or 3. Then *O* was shown the highest stimulus, and after it one of the remaining 13 stimuli, for which a judgment was required, then the lowest stimulus, and so on, until four ratings for each of the 13 stimuli had been obtained for both ascending and descending series.

RESULTS

The arithmetic mean of the four magnitude estimations of each stimulus made by each *O* was calculated; and the median of these means (between *O*s) is given in Table 1. There were small differences between the geometric means and the medians, making the latter the more advisable measure of central tendency. The intraindividual *SD*s were computed in the way described by Eisler (1962a) and are also given in Table 1. The arithmetic means of the category ratings calculated over presentations and *O*s are likewise given in Table 1. Throughout, ascending and descending series are treated separately.

Magnitude scales.—The magnitude scales for the ascending and the descending series were not entirely alike. Their curvilinearity in a double-logarithmic plot against (phys-

TABLE 1

STIMULUS VALUES (ϕ), SUBJECTIVE MAGNITUDES (ψ) WITH THEIR INTRAINDIVIDUAL *SD*s (σ), AND CATEGORY ESTIMATIONS (K) FOR ASCENDING AND DESCENDING SERIES OF VELOCITY

ϕ rpm	Ascending Series			Descending Series		
	ψ	σ	K	ψ	σ	K
20	1.650	.517	—	2.000	.450	—
25	2.375	.669	1.083	2.500	.910	1.042
36	4.375	.821	1.063	4.125	.943	1.063
45	5.375	.977	1.208	5.000	.986	1.500
55	7.500	.899	1.458	6.125	.945	1.646
60	8.000	1.287	1.521	7.375	2.257	1.729
70	9.375	1.776	2.021	9.000	1.383	1.938
82	10.688	1.522	1.958	10.125	1.866	2.042
90	13.000	2.187	2.125	11.500	1.604	2.229
113	17.500	2.356	2.646	14.875	2.044	2.542
126	18.750	2.531	2.667	16.875	1.885	2.646
153	23.125	3.715	2.896	19.750	3.639	2.979
170	27.375	2.747	2.958	22.500	2.986	2.938
189	31.875	4.396	2.958	24.875	4.429	3.000
200	34.250	3.160	—	30.625	3.506	—

ical) angular velocity showed that a threshold correction was needed if the psychophysical power law were to hold. The "subjective zero" (Ekman, 1958) was estimated from these log-log plots (cf. Mashhour, 1962) and straight lines were fitted to the equation

$$\log \psi = n \log (\phi - \phi_0) + C \quad [3]$$

(ϕ physical magnitude, rpm) by the method of least squares in steps of .1 rpm around the approximate ϕ_0 until a ϕ_0 value yielding the least sum of squared residuals was found.

The equations obtained in this way were

$$\log \psi_a = 1.09 \log (\phi - 7.4) - 0.981 \quad [4]$$

$$\log \psi_d = 1.11 \log (\phi - 2.0) - 1.096 \quad [5]$$

(Subscripts *a* and *d* refer to ascending and descending series, respectively.)

Figure 1 shows $\log \psi$ as a function of $\log (\phi - \phi_0)$. The difference in the exponents of the power functions for the ascending and the descending series (slopes of the straight lines in Fig. 1) is less than 2%, indicating that the exponent of the psychophysical function is invariant with the order of presentation of stimuli. The difference in the subjective range in log units for the two series is thus wholly due to the different correcting values ϕ_0. The small difference in C may indicate a minor difference in the subjective units for the two series.

Midpoints.—As Fig. 2 shows, the hysteresis effect, though small, is found also in the continuum of subjective velocity. In the plot, means of category ratings are plotted against physical angular velocity. The irregularity exhibited by the partition scale has been countered by taking moving averages over three successive points. This procedure is justified by the approximate linearity of the curves in the important region around the midpoint. The midpoints are considered to lie on the value "2" of the partition scale and are at 80 and 77 rpm for the ascending and descending series, respectively.

SDs of the magnitude estimates.—

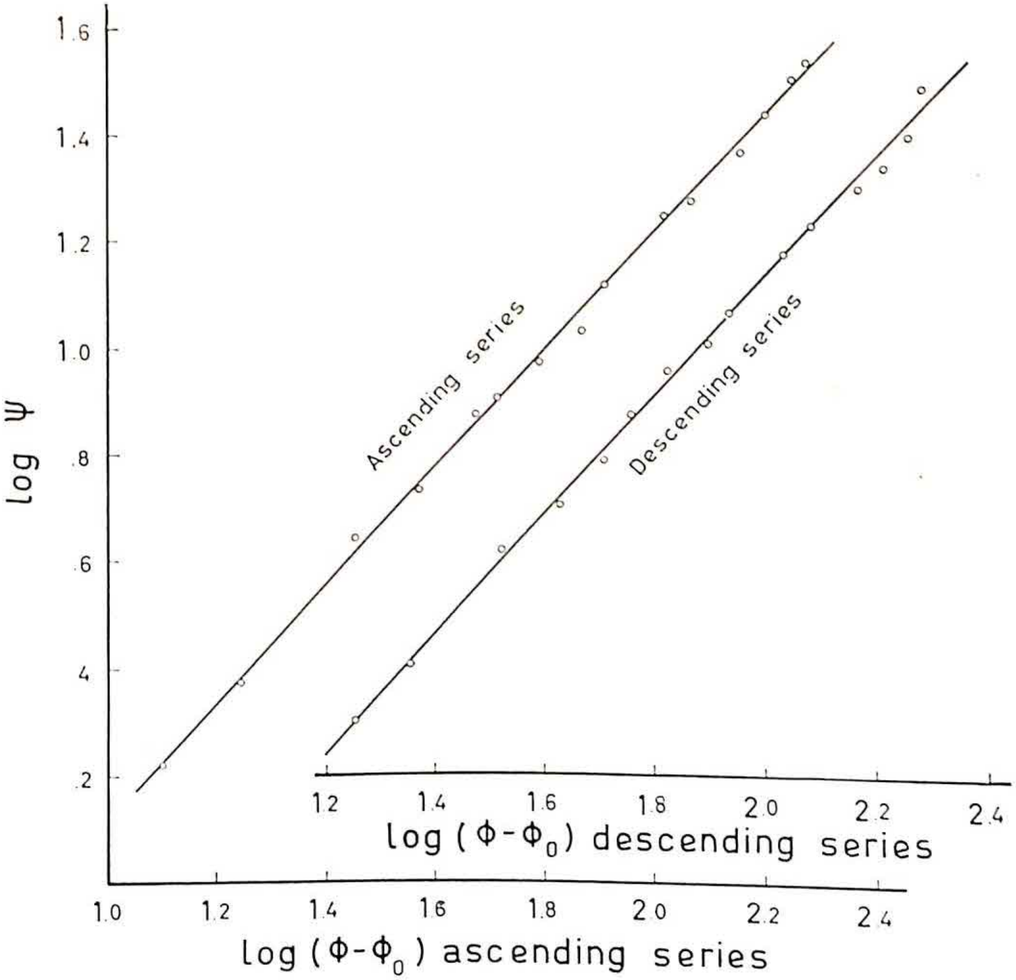


FIG. 1. Log magnitude estimations of subjective velocity as a function of the log of the difference in angular velocity (rpm) between the stimulus and the "subjective zero" (7.4 rpm for the ascending and 2.0 rpm for the descending series). (The straight lines are fitted by the method of least squares.)

The generalized form of Weber's law also holds for subjective velocity, as Fig. 3 shows. The straight lines describing the intraindividual *SD* as a function of subjective magnitude are fitted by the method of least squares. Note that unlike the experiments on loudness and softness (Eisler, 1962a) the two extreme points do not lie noticeably below the straight line as compared to the other points. In the experiments cited, *all* stimuli were shown in each presentation series (though in random order), so that extreme stimuli were probably more easily identified by *O*s, entailing spuriously low *SD*s; the procedure in the present experiment, in which only a random selection of stimuli was shown in each presentation series, seems to have prevented this effect.

For the ascending series, the slope of the line is .106 and the ψ intercept ψ_0, -4.266. The corresponding values for the descending series are .125 and -3.416.

Computation of the midpoints.—Equation 1 applied to the case at hand yields the following system of equa-

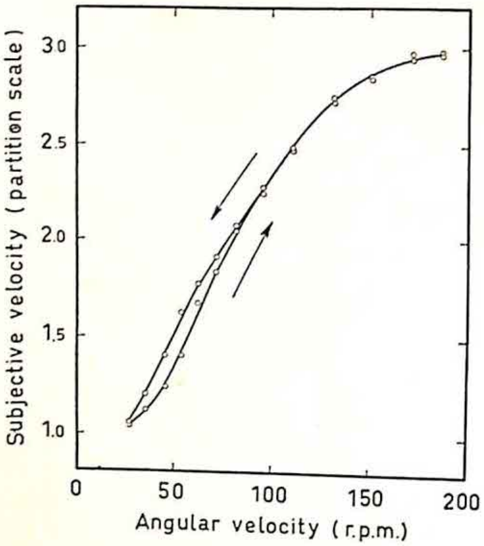


FIG. 2. Partition scale of subjective velocity as a function of angular velocity in rpm displaying hysteresis. (The circles are values smoothed by taking moving averages over three successive empirical pairs of values. The arrows indicate the order in which the stimuli were presented.)

tions:

$$3 = \alpha \log (\psi'_{15} - \psi_0) + \beta$$
$$1 = \alpha \log (\psi'_1 - \psi_0) + \beta,$$

where ψ'_1 and ψ'_{15} are the subjective magnitudes corresponding to the lowest and highest stimulus computed from Equations 4 and 5. After evaluation of α and β the midpoint in subjective units ψ_m is computed from

$$2 = \alpha \log (\psi_m - \psi_0) + \beta.$$

The midpoints calculated in this way are at 10.57 subjective units for the ascending series and 9.57 for the descending series. Figure 4 shows the partition scale as a function of the magnitude scale with the computed midpoints marked off as encircled dots. The circles on the curves represent values smoothed over three points, as in Fig. 2. The empirical midpoints as read off the curves are 11.1 and 9.6 subjective units.

Bearing in mind the many steps required in these computations, such a close agreement of the descending series is perhaps more surprising than the small discrepancy exhibited for the ascending series. The error, however, is big enough to reverse the

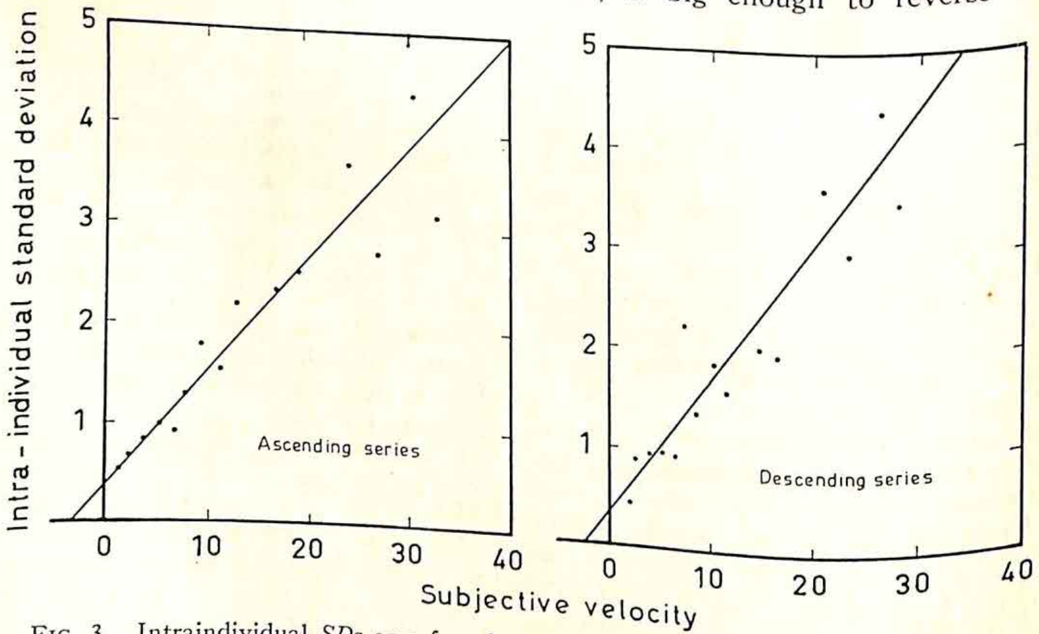


FIG. 3. Intraindividual *SD*s as a function of the magnitude scale of subjective velocity for ascending and descending series. (The *SD*s are computed according to Eisler, 1962a and are based on magnitude estimates calculated with Equations 4 and 5. The straight lines are fitted by the method of least squares.)

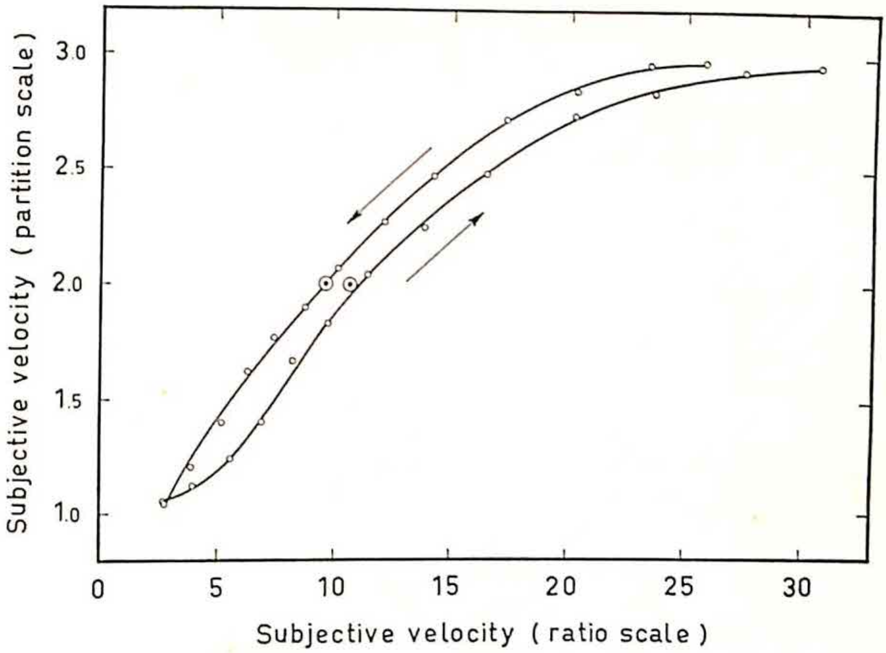


FIG. 4. Partition scale of subjective velocity as a function of the corresponding magnitude scale. (The same smoothing procedure as in Fig. 2 is used. The arrows indicate the order in which the stimuli were presented. The encircled dots are predicted bisection points.)

hysteresis effect if the computational step to the physical continuum is taken and the midpoints ϕ_m in rpm are calculated with Equations 4 and 5. For the descending series 77 rpm is computed in agreement with the outcome of the bisection experiment but for the ascending series 76 rpm is calculated and 80 rpm obtained empirically. Most probably, the error is due to the ψ intercept of the ascending series. A ψ intercept of -5.557 instead of -4.266 would have given an exact prediction.

DISCUSSION

The results of the experiments suggest that the model relating magnitude and category estimation scales also covers other types of partition scales. The hysteresis effect as expressed in different midpoints for different orders of presentation of stimuli, can be predicted from the magnitude scales. It is mostly, if not solely, due to different "subjective zeros" ϕ_0 in the psychophysical power function $\psi = k(\phi - \phi_0)^n$ for ascending and descending series. Whether the difference between the series in the corresponding subjective correction factor ψ_0 is a true difference is doubtful; its contribution to the hysteresis effect is therefore uncertain.

Precise knowledge of the zero point would prevent hysteresis. This seems to be the case in certain metathetic continua such as position (Stevens, 1957). It is also of interest to note that the pitch function changes with the anchoring of the zero (Stevens & Volkmann, 1940).

Another interesting aspect is the changes in dynamic range lying at the root of the hysteresis effect. The physical range is 1.0 log unit before the threshold corrections and rises to 1.18 and 1.04 log units for the ascending and descending series if expressed in terms of $\phi - \phi_0$. The corresponding values for the magnitude scales are 1.29 and 1.15 log units. This difference decreases again after the correction ψ_0 has been applied, giving .80 and .77 log units. It is this difference in range of .03 log units that produces the hysteresis.

The inflexion point in the ascending branch of Fig. 4 suggests that the partition scale taken as a whole is not quite consistent with Equation 1, which expresses the relation between category and magnitude scales. This is not entirely unexpected, since the procedure used here and the typical procedure employed to obtain category scales differ in several ways.

Naturally, it remains to investigate other subjective continua to see how general the description of hysteresis in terms of ϕ_0 is. At all events, the present findings indicate that the parameter ϕ_0 has not much to do with the absolute threshold, which probably lies far below 1 rpm (the angular velocity of a second hand of a clock). This is not so astonishing, since it may be requiring too much from an *O* that he keep in mind where his absolute threshold lies when making magnitude estimations. On the other hand, that *O*s under favorable conditions may be quite good at giving a meaningful value for ϕ_0 has been demonstrated by Stevens and Stevens (1960). A variation of ϕ_0 for luminance is described by Stevens (1962). In Stevens' case, however, the variation probably lies at the receptor level.

The present results may also throw some light on the problem of the time-order error, an effect that seems to be closely connected with hysteresis. If the parameter ϕ_0 takes different values depending on the order of presentation of a stimulus, the subjective magnitude pertaining to this stimulus will also be different—provided that the constant *k* in the power law does not change so that this effect is exactly canceled.

The same argument can be expressed in terms of Luce's psychophysical model (Luce, 1959, p. 43). Starting from the linear generalization of Weber's law

$$\Delta\phi = c\phi + d,$$

Luce shows that the exponent of the power function depends on the slope *c* and that the subjective zero ϕ_0 equals the ϕ intercept $-d/c$. It follows from the present experiments that order of presentation of stimuli does not affect the slope in Weber's law (invariant exponent) but does affect the constant *d*, or better, the ϕ intercept (different subjective zeros). A shift in ϕ_0 would thus be equivalent to a parallel translation of the straight line describing Weber's law, and would produce a change in $\Delta\phi$ entailing a displacement of the psychometric function: the time-order error.

We have not attempted to explain the shift of the subjective zero with the order of presentation of the stimuli. This is a topic of research of its own account and might be attacked for instance by investigating the parameters of Weber's law for different conditions.

REFERENCES

Eisler, H. Empirical test of a model relating magnitude and category scales. *Scand. J. Psychol.*, 1962, **3**, 88–96. (a)

Eisler, H. On the problem of category scales in psychophysics. *Scand. J. Psychol.*, 1962, **3**, 81–87. (b)

Eisler, H. How prothetic is the continuum of smell? *Scand. J. Psychol.*, 1963, **4**, 29–32. (a)

Eisler, H. Magnitude scales, category scales, and Fechnerian integration. *Psychol. Rev.*, 1963, **70**, 243–253. (b)

Ekman, G. Two generalized ratio scaling methods. *J. Psychol.*, 1958, **45**, 287–295.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Mashhour, M. On fitting some curvilinear functions to experimental data. Report No. 131, 1962, Stockholm University, Psychological Laboratory.

Stevens, J. C., & Stevens, S. S. Warmth and cold: Dynamics of sensory intensity. *J. exp. Psychol.*, 1960, **60**, 183–192.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

Stevens, S. S. To honor Fechner and repeal his law. *Science*, 1961, **133**, 80–86.

Stevens, S. S. The surprising simplicity of sensory metrics. *Amer. Psychologist*, 1962, **17**, 29–39.

Stevens, S. S., & Volkmann, J. The relation of pitch to frequency: A revised scale. *Amer. J. Psychol.*, 1940, **53**, 329–353.

(Early publication received January 29, 1963)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 537–545

EFFECTS OF PRIOR AND INTERPOLATED LEARNING ON RETENTION IN PIGEONS[1]

JACSUE KEHOE[2]

Brown University

A discrimination situation was developed for pigeons which would permit the investigation of retention problems in the animal laboratory. Changes in Retroactive and Proactive Inhibition (RI and PI) were studied as a function of the length of the retention interval (1, 10, or 30 days). A 5-key discrete-trial situation was employed in which a task was defined by the key color that was associated with reinforcement. Comparison of the number of correct responses made by the various groups on Day 1 of relearning permits the following conclusions: (a) Retroactive Inhibition was very prominent at all retention intervals. (b) The RI Ss relearning after a 30-day interval performed better than those relearning after a 1- or 10-day interval. (c) Ss learning an additional task prior to the task to be recalled made as many correct responses during the retention test as did control Ss. That is, no Proactive Inhibition was evident. (d) Both PI and control Ss relearning after a 30-day interval performed more poorly at recall than did those Ss relearning after a 1- or 10-day interval—a trend opposite to that shown by the Ss in the RI condition. A supplementary experiment showed that even when the number of prior tasks was increased from 1 to 2, no PI was evident after a 30-day interval—the only interval tested under this special interference condition.

The discrimination situation described below was developed to permit the systematic study of retention phenomena in the animal laboratory. Using pigeons as *S*s in wave length discrimination tasks, the following basic questions were explored. (*a*) Could Proactive Inhibition (PI) be demonstrated in the type of situation used here; more specifically, would the retention of a task be interfered with by learning that occurred prior to the acquisition of that task? (*b*) Could Retroactive Inhibition (RI) be demonstrated; that is, would the learning of an additional task during the retention interval interfere with recall? and (*c*) Would changes occur in these interference effects as a function of the length of the retention interval?

In their interpretations of these classic retention phenomena, investigators have often employed concepts originating from simple conditioning situations (e.g., Briggs, 1954). It has been assumed that, during the learning of a second task, the first-task response undergoes extinction (if the *S* is learning a new response to an old stimulus). It is this first-task extinction that has been used as a partial explanation of the phenomenon of Retroactive Inhibition. This extinguished first-task response is, moreover, presumed to recover in strength during the time that elapses between second-task learning and the recall test, paralleling the phenomenon of spontaneous recovery of an extinguished response in a simple conditioning situation.

---

[1] This article is based on a dissertation submitted to Brown University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. This investigation was carried out during the author's tenure as a Predoctoral National Science Foundation Fellow. The author gratefully acknowledges the assistance given to her by J. W. Kling throughout all phases of this research.

[2] Now a Public Health Service Postdoctoral Fellow in the Department of Experimental Psychology, Walter Reed Army Institute of Research, Washington 12, D. C.

How, then, would this extinction and later recovery of the first-task response affect *S*s in the various conditions of the traditional retention paradigms? For those *S*s who are to relearn their first task (RI *S*s), any recovery of this extinguished response should facilitate their recall performance. And, since an increase in the retention interval length permits more recovery of this response, the RI *S*s should perform better as the retention interval increases. Data that have been gathered relevant to the support of this hypothesis have been somewhat discrepant (see Slamecka & Ceraso, 1960) and therefore provide a rather unsatisfactory basis for predicting results under the present conditons.

For *S*s reinforced during relearning for performing the second-task response (PI *S*s), any first-task recovery would be interfering since these two responses presumably compete with one another. Consequently, as the retention interval increases and the first-task response is permitted greater recovery, PI should increase. In the few instances in which PI has been investigated as a function of the retention interval length, there does appear to be the predicted increase in the interfering effects of prior learning with time (see Underwood, 1957).

A decrement in recall shown by PI *S*s and an improvement shown by RI *S*s with time could also be predicted by assuming that the second-task response somehow grows weaker with time, without any absolute increase in the strength of the first-task, extinguished response. Consequently, it is necessary to include Control groups at all retention intervals if any support is to be derived for the hypothesized spontaneous recovery of the first-task response. For those *S*s learning only one task (Controls) there is no experimentally developed high strength response learned prior to their acquisition period. Consequently, there should be relatively little extinction necessary during learning, and, therefore, little subsequent recovery of incorrect responses. Since these *S*s should experience less interference from spontaneously recovering responses than those *S*s recalling the second of two tasks, less decrement with time would be expected from Control *S*s than from PI *S*s.

METHOD

Experimental Design

The effect of three retention intervals (1, 10, or 30 days) on relearning performance was investigated under each of three different interference conditions (Control, RI, and PI). The design of the experiment is represented in Table 1. The number in each cell refers to the number of *S*s participating in each condition. Before running all nine groups of the design simultaneously, some *S*s were assigned randomly to just the three Control conditions. Following these initial runs, *S*s were run in a series of replications of all nine conditions of the design. This two-stage process resulted in a larger number of *S*s in the Control conditions.

Subjects

The *S*s were White Carneau cocks ranging from 4 to 7 yr. in age. Of the 80 pigeons

TABLE 1

EXPERIMENTAL DESIGN

Interference Condition	Retention Interval (Days)		
	1	10	30
Control	10	9	11
PI	4	4	5
RI	4	5	5

Note.—Number within each cell indicates *N* per group.

who were exposed to the discrimination training conditions, 23 were dropped for the following reasons: 17 either failed to transfer from the pretraining procedure to the discrimination problems, or extinguished during the early acquisition of the second task; 2 were dropped due to *E* error in programing the conditions; 2 died during their retention interval; and 2 were dropped prior to recall testing for showing an excessive number of between-day decrements in performance which would have rendered their recall data uninterpretable.

Each bird was maintained throughout the experiment at 75% of his maximum body weight. Purina pigeon chow was used both for the maintaining diet and as the reinforcer during experimentation.

Apparatus

The compartment in which the bird was placed for experimentation was a blackened aluminum and Plexiglas box with five Gerbrands-type response keys mounted on the front wall of the box. The keys were differentially illuminated by two systems of light sources. The first, used only during pretraining, consisted of five white 7-w. 110V ac bulbs which were wired and placed so as to illuminate each of the keys with any one of the five colors. For this purpose 7-w. 110V ac bulbs were used in conjunction with five types of gelatin filters from the Wratten narrow-band series (No. 70, 72B, 73, 74, and 75). The system was wired so that (*a*) no two keys were lighted with the same color during any given display, and (*b*) each key was lighted with each of the five colors equally often within a session.

The aluminum box and the stimulus system were enclosed in a ventilated refrigerator shell which permitted observation of the bird, but was adequately shielded from all light external to the compartment. A speaker placed inside the refrigerator shell provided white noise which masked apparatus sounds.

A response was defined as a break in the normally closed contacts of any one of the five keys. Each break was recorded independently on Elmeg printing counters in such a way that it could be identified in terms of the position of the key pecked, the color currently illuminating each of the keys, and the order in which that particular response occurred with respect to all other responses made in the session.

Procedure

Pretraining.—The pretraining involved three stages: (*a*) magazine training, (*b*) shaping and training of the pecking response, and (*c*) a session providing practice in pecking at each of the five key positions.

During training of the pecking response only the center key was exposed, and white light was used to illuminate the key. Each response produced a 4-sec. exposure to food, followed by a 2-sec. blackout condition. One hundred such trials were given.

In the last pretraining session, the position of the white light was changed between each presentation, appearing equally often on each key throughout the session. During the first 50 stimulus exposures of this training period, a response to the lighted key was followed by a 3-sec. food exposure and a 2-sec. blackout period. For the last 50 exposures, the blackout was lengthened to 27 sec.—the duration to be used during discrimination training.

Discrimination training.—There were two tasks used in this experiment. One of the two was defined by the reinforcement of pecks to the key illuminated with green light (Filter No. 74), regardless of its position, while the other was defined by reinforcement of pecks to the red key (Filter No. 70). Throughout the investigation the only factor differentiating one task from the other was this reinforcement contingency. The task the Control *S*s were to learn and the order of tasks for the RI and PI *S*s were assigned at random.

Throughout discrimination training each daily session consisted of 50 exposures of the stimulus displays. The stimuli were turned on after the bird was placed in the dark compartment and following his response to one of the five keys the stimulus lights went out. If the response had been the correct one, it was followed by a 3-sec. exposure to food and a 27-sec. blackout. If it were incorrect, no reinforcement was given and the response was immediately followed by a 30-sec. blackout. Only one peck was required to initiate this blackout or reinforcement-blackout condition. The display was changed automatically following each response.

Between the twentieth and twenty-first and fortieth and forty-first presentations of the stimuli a so-called correction trial was introduced. On such trials, only the center key was lighted and the color currently associated with reinforcement illuminated the key. This procedure, since it virtually forced the bird into responding to the correct color and thereby receiving the reinforcement,

TABLE 2

Mean Days to Criterion and Mean Correct Responses on Criterion Days

Experimental Condition	Mean Number of Days to Criterion		Mean Number Correct on Criterion Day	
	Task 1	Task 2	Task 1	Task 2
Control				
1 day	3.0	—	46.8	—
10 day	3.0	—	48.33	—
30 day	3.64	—	46.73	—
RI				
1 day	3.5	4.0	47.75	48.0
10 day	3.6	4.2	47.2	48.2
30 day	3.4	4.2	47.4	47.8
PI				
1 day	3.5	4.5	48.25	47.75
10 day	3.5	4.25	47.0	46.0
30 day	3.8	5.0	47.0	47.6
H	3.17	5.10	9.92	8.26
df	8	5	8	5
p	<.95 >.90	<.50 >.30	<.30 >.20	<.20 >.10

helped avoid extinction during the early acquisition of both tasks.

The birds were run one session per day until they reached a criterion of 45 or more correct responses out of the possible 50 within a session. The two correction trials were not included in this calculation. Control animals began their retention interval after reaching criterion performance on their only task. Experimental animals, on the day following criterion performance on their first task, were put in under the new reinforcement contingency defining their second task. They were run through a similar series of daily sessions until they met criterion on their second task, at which time their retention interval commenced.

Relearning.—Relearning sessions involved the use of the same basic procedures as the acquisition sessions. The control animals, who had been reinforced for only one of the responses, were run under conditions identical to the acquisition sessions. For the PI birds, the second-task response was reinforced during relearning, while for the RI birds, reinforcement was delivered upon occurrence of the first-learned response.

Results

Acquisition Data

The mean number of days required by each group of *S*s to meet the criterion of 45/50 responses correct is presented in Table 2. This table also lists the mean number of correct responses made by these groups on the day of criterion performance on their respective tasks.

All the means within a column represent scores that were obtained under similar conditions. The *p* value listed at the bottom of each column indicates the probability of the chance occurrence of the mean differences evident within that column, and thereby gives an evaluation of pretreatment differences. As one can see by observing the data and their statistical evaluation (Kruskal-Wallis

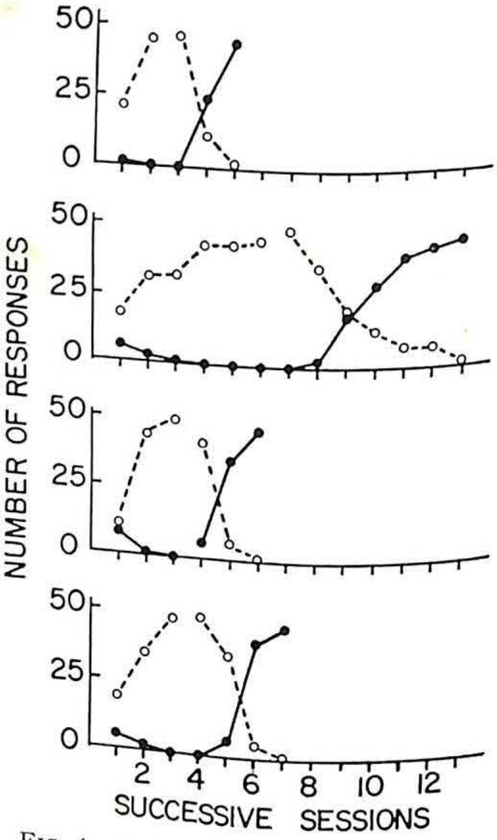


Fig. 1. Individual acquisition curves for all *S*s in the 10-day PI group. (The dashed line refers to the frequency of the response reinforced during Task 1; the solid line, to the Task 2 response. The break in each set of curves indicates the change from Task 1 to Task 2.)

analysis of variance), groups that have been treated similarly perform similarly.

There are reliable differences, however, between Task 1 and Task 2 performance for all groups learning more than one task. The mean number of days required to learn the second task was greater than that required to learn the first task. Eighteen of the 27 *S*s required more days to meet the second task criterion, and only 2 of the 27 *S*s spent fewer days on Task 2 ($p < .002$). This negative transfer is not surprising considering the fact that an identical-stimulus—changed-response design was used. The mean number of correct responses on the last day did not differ between tasks ($p = .8$), which was, of course, the result aimed for when placing the criterion restriction on the *S*'s performance.

Sample data for the two-task groups can be seen in Fig. 1, which presents the acquisition scores for both Tasks 1 and 2 for all birds in the PI, 10-day group. The solid line indicates the frequency of occurrence of the response that was reinforced during the second task; the dashed line, the frequency of the first-task response. During second-task acquisition, the extinction of the first-task response results in its decrease in frequency, and the reinforcement of the new response results in its gain in strength. Data obtained from Control animals are similar to the first-task data illustrated here.

TABLE 3

MEAN CORRECT RESPONSES AT RECALL AND MEAN DAYS TO RELEARN

Measure	Condition	Retention Interval (Days)			Statistical Evaluation
		1	10	30	*H*
No. correct Rs at recall	Control	48.6	46.9	42.3	9.55**
	PI	47.5	48.7	39.6	7.32*
	RI	4.5	5.8	15.8	5.52
No. days to relearn	Control	1.1	1.1	1.7	
	PI	1.0	1.0	2.0	
	RI	4.5	4.8	3.3	

* $p < .05$.
** $p < .01$.

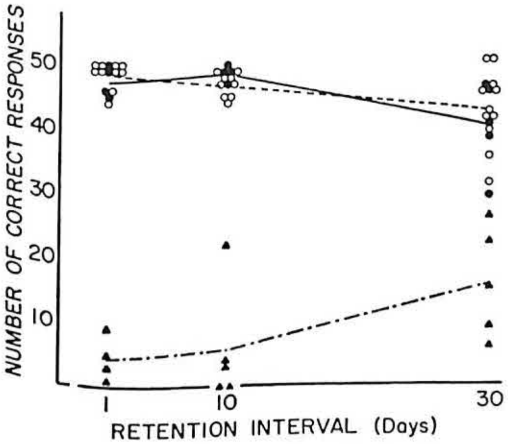


FIG. 2. Number of correct responses made by each *S* during his recall session. (Scores of Control *S*s are represented by open circles and the dashed line; those of PI *S*s, by closed circles and the solid line; and the data of the RI *S*s, by triangles and the dash-dot line.)

Day 1 Relearning: Number of Correct Responses

Figure 2 presents the number of correct responses made by each *S* on the first relearning day, and Table 3 presents the mean performance of the various groups.

Proactive Inhibition.—When the number of correct responses made by the members of the PI and Control groups is compared at each of the retention intervals, extreme overlap is observed and no difference between the performance under the two interference conditions is evident ($F < 1$). There is no indication that the *S*s who learned an additional task prior to the acquisition of the task to be recalled made fewer (or more) correct responses during the Day 1 Relearning session than did Control *S*s who learned no additional task. That is,

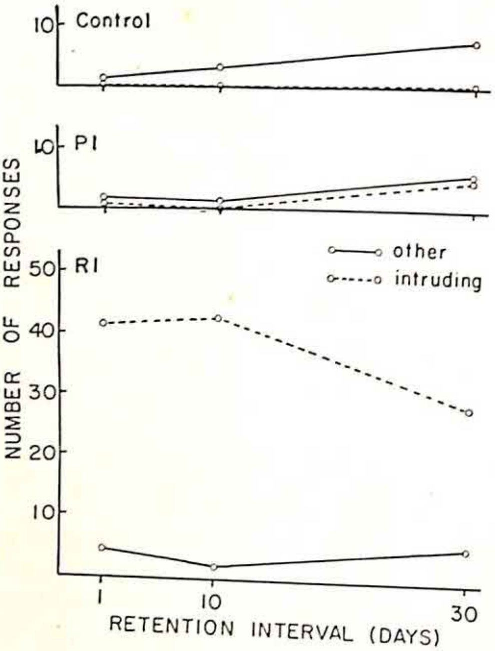


FIG. 3. Mean number of "other" responses (solid line) and "intruding" responses (dashed line) made by the Control, PI, and RI groups.

no PI was observed, using the traditional measure of such interference effects.

Under both PI and Control conditions, the retention interval was a significant determiner of performance. Increasing the interval from 1 to 10 days had no apparent effect on behavior, but when the interval was 30 days in length, a decrement in retention resulted (see Table 3). For the Control group, at least, the variability of Day 1 Relearning performance also seemed to change with time. The 30-day group not only exhibited a lower mean performance than the 1- and 10-day groups, but the scores in that group were also much more heterogeneous. This may, of course, be merely an artifact due to the limiting 100% level of performance at which many of the members of the 1- and 10-day groups were working.

Retroactive Inhibition.—A comparison of the Day 1 Relearning scores from the Control *S*s and those of the RI groups leaves no doubt that the learning of a second task had a dramatically detrimental effect on the retention of the first-learned task ($F = 460$; see Table 3 and Fig. 2). In contrast to the trend shown as a function of time by both the PI and Control groups, under the RI condition an increase in the retention interval appears to have resulted in an increase in the number of correct responses at recall.

Day 1 Relearning: Analysis of Errors

In addition to a comparison of the number of correct responses made by the various groups, it is possible to compare the types of errors made during the retention test and perhaps thereby to reveal differences as a function of experimental history that would not be apparent by observing only the frequency of occurrence of the correct response. In order to facilitate such a comparison, the following categories of response have been defined. Only two of the five responses were ever associated with reinforcement. Of these two "task" responses, the one currently being reinforced defines the *correct response.* An *intruding response* is defined as the task response not currently being reinforced. This categorization is used even for Control *S*s who learned only one task, for by separating these "intrusions" from the other errors of the Control *S*s, it is possible to obtain a baseline frequency of occurrence against which to measure the intertask interference shown by animals for whom these responses can meaningfully be labeled as intrusions from other learning. *Other responses* are those made to the three keys that were never associated with reinforcement.

Figure 3 presents the error data for all nine groups of the experiment.

Proactive Inhibition.—Although no PI was evident by standard measures, observation of the error data does suggest that the experimental history of the *S*s in the PI groups did affect their retention test performance. The Control *S*s made essentially no intruding responses at any of the time intervals. Similarly, the PI *S*s who recalled after either 1 or 10 days showed practically no intrusions. However, *S*s in the 30-day PI group made almost as many intrusive errors as they made "other" responses.

Retroactive Inhibition.—The number of intruding responses made by the RI groups is far greater than that seen in the Control or PI data. As the retention interval increased in length, the number of intruding responses decreased, while the number of correct responses increased. The number of "other" responses remained at approximately the same frequency at all retention intervals.

Relearning: Number of Days to Relearn

The number of days to relearn is negatively correlated with the number of correct responses on Day 1 Relearning (see Table 3). This is to some extent unavoidable, since if *S* performed at criterion level or above on the first relearning day, he received no more training. Only *S*s who made fewer than 45 correct responses were given more than one relearning session.

As Table 3 shows, the RI *S*s required many more days to relearn than did the PI or Control birds, and, in contrast with the PI and Control *S*s, they showed a decreased number of days to relearn as a function of time.

DISCUSSION

The experiment was designed to serve two purposes: first, to demonstrate the frequently observed phenomena of RI and PI in this new experimental situation; and, secondly, to investigate the changes in the amount of these interference effects with time.

The first of these purposes was only partially served. Although RI was dramatically evident at all retention intervals, there was no evidence of PI as measured by comparing the number of correct responses made by the PI and Control groups during the retention test. The PI 30-day group did show a greater preponderance of intrusive errors than did the comparable Control group, but the interference was certainly not of the magnitude expected from the design employed.

The effects of varying the retention interval length—the second question asked in this study—were also somewhat different from those anticipated. Although the empirical basis for prediction is somewhat unclear, most investigators have predicted increased PI and decreased RI as a function of time (e.g., Estes, 1959; Underwood, 1948). It has been hypothesized that the first-task response, extinguished during second-task learning, would spontaneously recover during the retention interval, and that the amount of recovery would be a direct function of time. This recovery would improve the recall of the RI *S*s and hinder the performance of the PI *S*s. The PI *S*s did show the predicted decrease in performance and the RI *S*s the predicted increase. However, since the Control and PI *S*s showed an equal decrement with time, there is no differential support given for the hypothesis that the changes occurring in the two task groups were due to a recovery in first-task strength.

The most surprising result of the experiment is the minimal effect that prior learning had on recall. PI has been shown frequently in the verbal learning situation (see Underwood, 1957), yet seems to be quite questionable here. A number of differences in the two approaches might explain the disparate results. For example, in the verbal learning task, the presentation of the stimulus is always followed in a few seconds by a presentation of the correct word or syllable, independent of the *S*'s behavior. Consequently, if he makes the proper response, he is "reinforced" by its presentation, and if he makes an error he is "corrected." In the situation used in this study, on the other hand, no indication is given following an error as to which response should have been made, and *S* is not forced into making the correct one. It is not unreasonable to assume, then, that the old response would extinguish differently in the two situations. That is, in the verbal learning task, the new response is probably learned relatively more quickly, gaining in strength over the old response prior to its complete extinction. In this situation, in contrast, the old response must be quite well extinguished before the new response even occurs. Whether or not the use of a correction procedure would have increased the amount of PI is a question that awaits an experimental answer.

This situation also differs from the verbal learning task in terms of the number of extinction sessions that the response undergoes prior to the recall test. With the procedures employed here, there are many sessions during which the first-learned response undergoes extinction. These sessions are separated by a 24-hr. interval, and it is likely that spontaneous recovery of the extinguished response occurs during this time. The response would then undergo a repeated extinction in the following session. Such repeated extinctions probably reduce the amount of recovery that will occur during the retention interval, much in the same way that repeated extinctions of a single response have been shown to reduce greatly the amount of recovery that will occur following the last extinction session. In the verbal learning situation, where learning criterion is met in one session, this repeated recovery-extinction process would not be as prominent, therefore greater first-task recovery would be likely during the retention interval.

A similar repeated-extinction hypothesis has been suggested as a possible explanation for superior retention following acquisition under distributed rather than massed practice. In interpreting such a finding obtained under a similar-stimulus—changed-response design, Underwood and Schulz (1961) suggested that:

> If a recovery process is assumed to occur over time, under distributed practice the extinguished associations will recover more between each trial than they will under massed practice. Such recovered associations would have to be extinguished again. It may be assumed that this alternating recovery and extinction process will lead to a more permanent extinction of the interference tendencies thus facilitating recall following distributed practice (p. 234).

Before modifying the procedures in line with the hypotheses suggested above, a further attempt was made to demonstrate PI in this situation without altering the basic nature of the task. The following study was undertaken to see if PI could be demonstrated by simply increasing the number of prior tasks from one to two, since similar manipulations have been shown to be an effective way of increasing the amount of PI in the verbal learning situation (see Underwood, 1957).

Two groups of PI *S*s were run in this supplementary experiment. One of the groups ($N = 7$) learned only one prior task; the other ($N = 8$) learned two prior tasks. All *S*s in this experiment recalled their last-learned task after a 30-day retention interval. Reinforcement of pecks to the key illuminated with the blue-green light (Filter No. 75) defined the third task. All possible combinations and orders of tasks were represented in the design. With the exceptions implied in these design changes, the procedures were the same as those used in the main investigation.

By comparing the Day 1 Relearning performance of the 30-day Control *S*s with the performance of the 30-day two- and three-task PI *S*s it is possible to obtain an estimate of the effect on recall of manipulating the amount of prior learning. Since the difference between the means of the two-task groups from the main experiment and this supplementary study did not reach significance ($p = .25$), the data from the two experiments were combined for this comparison. The mean number of correct responses on the first relearning day is essentially the same for all three groups: 42.27, 42.07, and 42.00 for the *S*s learning one, two, and three tasks, respectively. Although the mean performance does not appear to be affected, there does seem to be a reduction in the variability of Day 1, Relearning scores as the number of prior tasks increases from zero to two (*SD*s = 6.5, 4.82, and 3.02 for the one-, two-, and three-task groups, respectively).

In summary, it appears that the differences between the data arising from this situation and those typically obtained in human studies cannot be resolved by simply increasing the amount of prior learning. Perhaps if some changes were made in the basic structure of the task (such as those alterations suggested above) the data might show closer resemblance to the verbal learning data. If such proved to be the case, some new knowledge might be obtained concerning the variables affecting interference in retention.

REFERENCES

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. II. McGraw-Hill, 1959. Pp. 380–491.

SLAMECKA, N. J., & CERASO, J. Retroactive and proactive inhibition of verbal learning. *Psychol. Bull.*, 1960, **57**, 449–475.

UNDERWOOD, B. J. "Spontaneous recovery" of verbal associations. *J. exp. Psychol.*, 1948, **38**, 429–439.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

UNDERWOOD, B. J., & SCHULZ, R. W. Studies of distributed practice: XX. Sources of interference associated with differences in learning and retention. *J. exp. Psychol.*, 1961, **61**, 228–235.

(Received July 16, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 546–551

SUCCESSIVE ACQUISITIONS AND EXTINCTIONS IN A T MAZE[1]

JOHN W. COTTON AND GLEN D. JENSEN

University of California, Santa Barbara *DePaul University*

80 male albino rats received 20 trials of noncorrection training in a single unit T maze. Then randomly assigned groups of 20 Ss received 2 sessions of training including from zero to 2 acquisition sessions and the remainder of extinction sessions (in all 4 possible orders) after which all groups received an additional acquisition session and than a session of extinction. No evidence was found of inferior performance on Days 4 and 5 for groups receiving previous extinction trials. When groups were compared on trials for which previous numbers of reward were equated, there was some indication of superiority for groups having extinction trials on Days 2 and/or 3.

This experiment was designed to test the hypothesis that T maze performance on reacquisition trials following varied numbers of extinction trials would, except for a transition phase when reward is first reintroduced, be essentially equivalent to that obtained if those extinction trials had been acquisition trials. This hypothesis is a consequence of the assumption that habit strength grows on all trials, not just rewarded trials (Spence, 1956, p. 94).

Four groups of rats were differentiated by the pattern of daily rewarded (A) and nonrewarded (E) sessions that they received. The patterns were AAAAE, AEAAE, AAEAE, and AEEAE, where the first letter denotes the schedule obtaining on Day 1, the second letter the schedule for Day 2, and so on. Under the presumption that acquisition of speeds and correct turns is not complete before the end of 4 days under the procedure used, this design allows for 12 tests of the general hypothesis.

Since the experiment was performed, a modified Hullian theory has been developed which makes quantitative predictions regarding experiments such as this (Cotton, Jensen, & Lewis, 1962).

METHOD

Subjects.—The *S*s whose data are reported were 80 naive male albino rats of the Sprague-Dawley strain from 52 to 74 days in age with a mean of 65 days. Their mean weight was approximately 187 gm. In addition to the 80 rats which completed the experiment, 5 were discarded because of failure to learn, and 6 because of procedural error. The *S*s were assigned at random to four groups, subject to the restriction that the first 4 *S*s run included one member of each group, the second 4 included one member of each group, and so on. Since the decision to include the final extinction day was made after the experiment was begun, the respective groups (as listed in the second paragraph of this article) have *N*s of 13, 12, 14, and 15 on the last day but 20 on all other days.

Apparatus.—The "narrow" T maze used by Cotton, Lewis, and Metzger (1958) was employed with a Standard Electric timer reading in .01-sec. units.

Procedure.—Prior to the first day of experimental treatment each *S* was placed on a deprivation and taming schedule for 7 days. During each of these days the *S* was presented

---

[1] This experiment was performed and largely analyzed at Northwestern University under National Science Foundation Grants G-2672 and G-8706. The authors gratefully acknowledge the influence of Donald T. Campbell upon the treatment of incentive magnitude given herein.

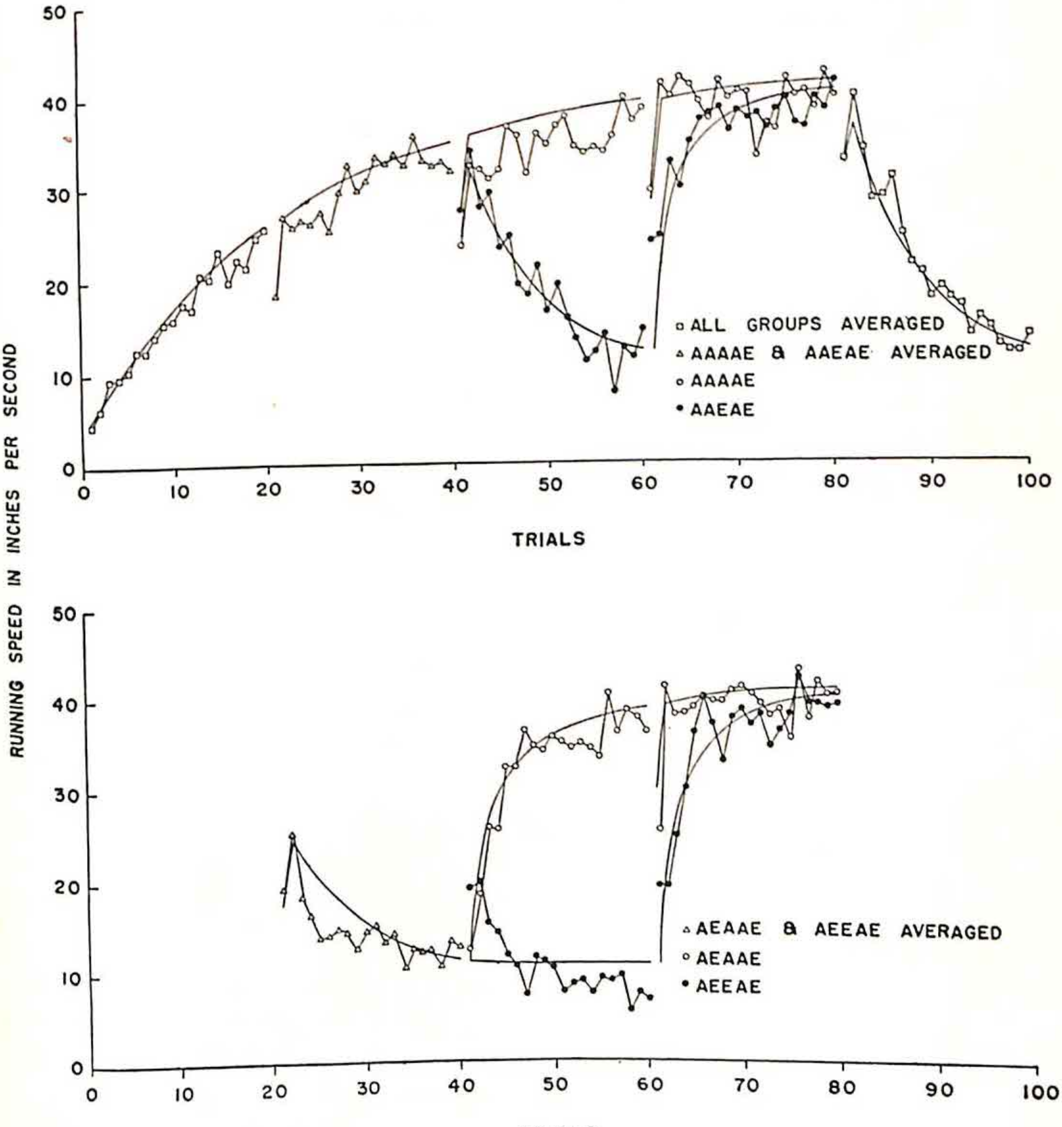


FIG. 1. Empirical and theoretical mean speeds in all stages of the experiment.

with 10 gm. of finely ground Purina lab chow in a tin cup inside its home cage at 1 hr. past the time the *S* was later to be run. It was not necessary to remove the food once presented because all *S*s were able to finish it within 30 min. after the first day or two of scheduling. Thus the *S*s spent 23–23.5 hr. of the day deprived of food. During each of these 7 days the *S* was handled and tamed for 5 min. and was encouraged to eat six 45-mg. "reward" pellets from a porcelain dish. Any pellets not eaten during the 5 min. of taming were placed in *S*'s home cage.

Twenty-three hours after the seventh presentation of food, each *S* received 20 acquisition trials in the maze. On each trial *S* was placed in the start box and allowed to orient toward the door giving access to the stem of the maze. Then *E* opened the door, starting the timer. This timer stopped when *S* reached a point 46.25 in. down the stem. When *S* entered an arm, the corresponding arm door was closed immediately, preventing retracing. The *S* was left in the arm for a minimum of 10 sec. but was allowed more time if it was a "correct" choice and *S* had not eaten the reward pellet by the end of 10 sec. The "correct" side for each *S* was determined according to a prearranged random schedule, subject to the restriction that half of the *S*s in each group learned to run to each side. If *S* failed to choose an arm within 2 min., he was manually lifted and placed in the correct goal box and allowed to eat the pellet. The intertrial interval was 15 sec. and was spent in the *S*'s home cage with free access to water. All further acquisition sessions were carried out in identical fashion.

Extinction sessions differed from acquisition sessions only in that food (and food dish) was never present and in that *S* was placed in a goal box for 10 sec. if he failed to choose in less than 2 min. *Which* goal box he was placed in was determined by means of two counterbalancing sequences, RLLR and LRRL. One of these sequences was assigned to each rat before each extinction session by means of a coin toss.

All running time scores were converted to their reciprocals because reciprocal running time scores are known to be more nearly normally distributed than raw running time scores. After multiplication by 46.25 these reciprocals became speed scores in in/sec.

RESULTS AND DISCUSSION

The unconnected data points in Fig. 1 show the course of running speed as a function of trials. An analysis of variance comparing the four groups and using the mean speed of each *S* during all 20 trials of Day 1 yielded nonsignificant results; $F(3, 76) = .67$, $p > .50$. The unconnected data points in Fig. 2 show the course of choice behavior. Blocks of five trials have been averaged and plotted at the abscissa points which represent the midpoints of five-trial blocks. An analysis of variance using the percent correct scores for all trials of Day 1 (which were approximately normally distributed) failed to yield significant results; $F(3, 76) = 2.19$, $p > .05$.

Since the groups did not differ significantly on either response measure at the beginning of training, several stringent tests of the general hypothesis could be made. They are summarized in Table 1. Note that the first two comparisons of running speeds indicate that habit strength increased during extinction, though the second statistic did not quite reach the .05 level of significance. The third comparison of running speeds did not even approach statistical significance, but this is likely due to the fact that the AAEA animals were running at a nearly asymptotic speed when the first extinction experience began.

TABLE 1

TESTS OF THE NULL HYPOTHESIS THAT EXTINCTION TRIALS DO NOT FACILITATE REACQUISITION PERFORMANCE

Data for Comparison	Test Employed	*df*	Value Obtained
Trials 76–80 for AEEA vs. Trials 36–40 for AAAA and AAEA (speeds)	Analysis of variance	1, 58	$F = 6.37^*$
Same as above except that choice data were used	Kruskal-Wallis *H* test	1	$H = 3.04$
Trials 56–60 for AEAA vs. Trials 36–40 for AAAA and AAEA (speeds)	Analysis of variance	1, 58	$F = 3.64$
Same as above except that choice data were used	Kruskal-Wallis *H* test	1	$H = .66$
Trials 76–80 for AEAA and AAEA vs. Trials 56–60 for AAAA (speeds)	Analysis of variance	1, 58	$F = 1.39$
Same as above except that choice data were used	Kruskal-Wallis *H* test	1	$H = 2.3$

* $p < .05$.

Though the choice data are graphically in line with the general hypothesis in every instance, statistical significance was not obtained in any case. While these results do not support the general hypothesis, neither do they mitigate against it in an important way: as Fig. 2 shows, the animals approached asymptotic performance on this response measure very quickly, thereby minimizing the possibility of obtaining significant differences on the comparisons.

The tests which have already been described are all such that rejection of the null hypothesis is consistent with the assumption that habit

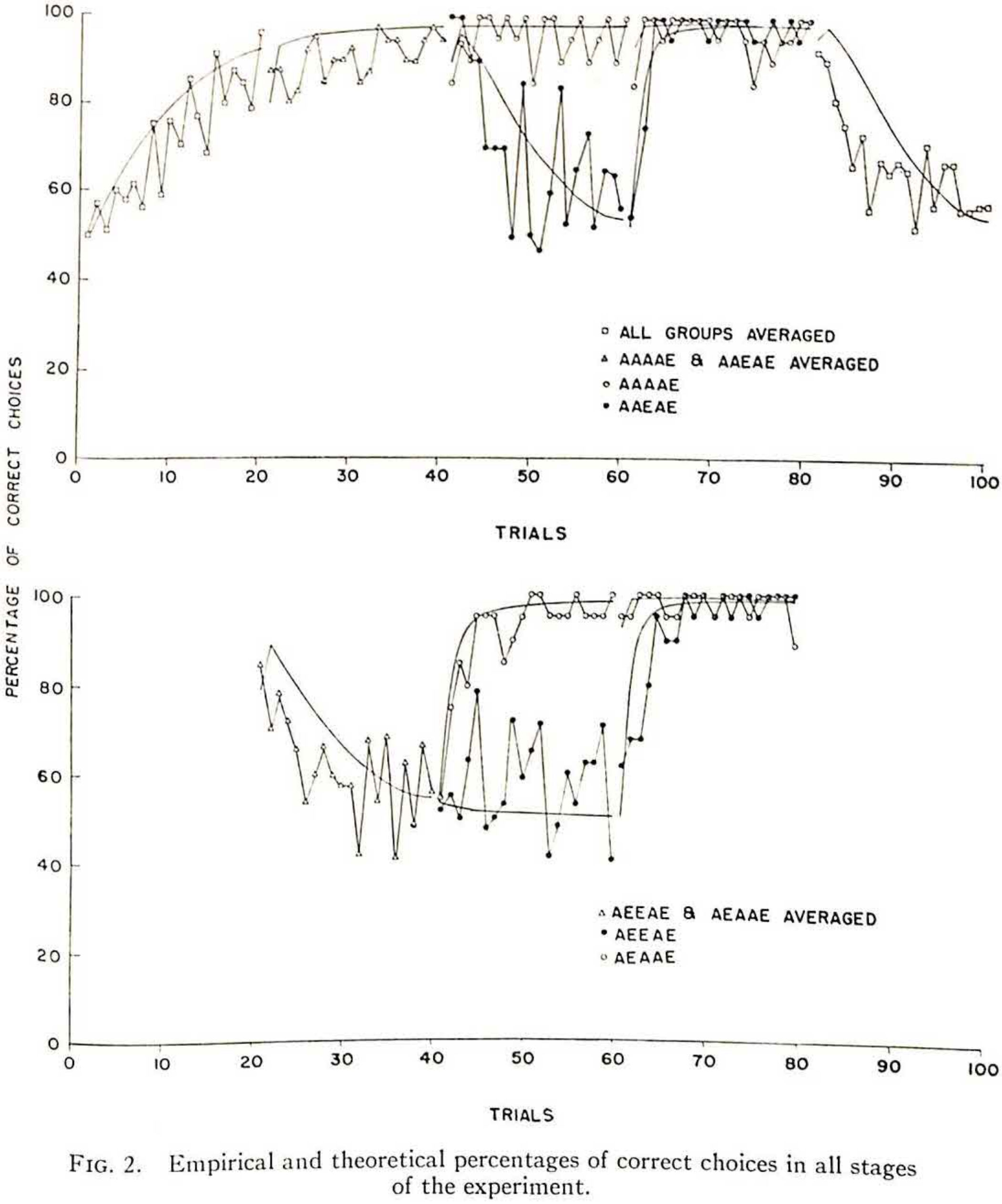


FIG. 2. Empirical and theoretical percentages of correct choices in all stages of the experiment.

strength increases during extinction. Other possible tests are such that acceptance of the null hypothesis is consistent with the assumption that habit strength increases during extinction. The general hypothesis implies that groups will not differ appreciably at the end of any session which is a reacquisition session for some groups and continued acquisition for one or more other groups. The AEAA rats did not differ significantly from the AAAA animals on the last five trials of Day 3 with respect to speeds or choices. The four groups did not differ significantly at the end of Day 4 or during all 20 trials of Day 5 with respect to speeds or choices. Thus, all six tests of this variety support the general hypothesis that

TABLE 2

PARAMETER VALUES USED IN FIG. 1 AND 2

Parameters	Values
B	.923
i	.01725
A_A	5.18
K_o	43.29
j	.479
A_E	10.64
k	.0672
C	.418
m	.0561
A_R[a]	
Group AAEA	11.89
Group AEAA	11.61
Group AEEA	10.68

[a] Strictly speaking, the A_R values are not parameter values but are consequences of the theory. They are presented to facilitate calculation of the theoretical curves.

habit strength increases during extinction.

Curve-fitting the data.—The smooth curves in Fig. 1 result from fitting the equations of the Cotton, Jensen, and Lewis (1962) paper to the present running speed data. The basic equation used was:

$$R_S = A + KHR \qquad [1]$$

where R_S is mean running speed, A is the lowest speed attainable in a given stage of the study, K is *incentive motivation* (which increases during acquisition and declines during extinction), H is *habit strength*, and R is a construct called *retention percentage*. Because incorrect choices were never rewarded by food in the present study, the equations in the previous article were simplified to omit separate reference to K values for correct and incorrect turns:

$$K = (K_O - A_A)10^{-j/N_S} \qquad [2]$$

during the first acquisition, where K_O is the maximum K value attainable with the reward magnitude employed, A_A is the A value for the first acquisition period, j is a growth parameter, and N_S is the number of trials under the present reinforcement schedule. During extinction

$$K = K'(10^{-kN_S}) \qquad [3]$$

where k is a decline parameter and

$$K' = \frac{KH + A_A - A_E}{H} \qquad [4]$$

with the K and A_E of Equation 4 being the value extrapolated for the first extinction trial from Equation 2, and the A for extinction, respectively.

During reacquisition, K has the same properties as in Equation 1 except that A_R (the A for reacquisition) replaces A_A. The habit strength equation is:

$$H = 1 - B(10^{-iN}) \qquad [5]$$

where B and i are curve-fitting constants and N is the number of previous trials, including reinforced and nonreinforced trials.

Note that the predictions in Fig. 1 are good except for Days 2 and 3 of Group AEEA. This suggests that the value of k should be larger if 20 acquisition trials precede extinction than if 40 do. This modification in k will not be made yet, pending verification of its necessity in related experiments.

Table 2 presents all parameter values for this experiment. For habit strength and retention they are the same as in the previous study.

The smooth curves in Fig. 2 present percentages of correct responses based on the theory previously used. The probability of a correct response is 81/80 of the normal curve integral from $z = -2.5$ to $KHR/\sqrt{2}\sigma$ where

$$z = \frac{\dot{\bar{E}}_C - \dot{\bar{E}}_I - KHR}{\sqrt{2}\sigma} \qquad [6]$$

and $\dot{\bar{E}}_C - \dot{\bar{E}}_I$ is normally distributed with mean KH but truncated at $\pm 2.5\ (\sqrt{2}\sigma)$ from its mean. $\dot{\bar{E}}_C$ and $\dot{\bar{E}}_I$ are the momentary effective reaction potentials for the correct and incorrect turns, respectively. All empirical parameters for fitting the choice data were obtained from the speed data, as in Cotton, Jensen, and Lewis. Irregularities in the choice data are partially understood by noting that for a group of 20 *S*s making correct and incorrect choices randomly on a trial with R_P (the probability of a correct response) equal to .90, the standard error of the proportion is .067. This rises to .112 for R_P = .50.

REFERENCES

Cotton, J. W., Jensen, G. D., & Lewis, D. J. Wide variation in spontaneous recovery interval as an influence on reacquisition of T maze behavior. *J. exp. Psychol.*, 1962, **63**, 555–562.

Cotton, J. W., Lewis, D. J., & Metzger, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

(Received July 18, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 552–558

SEMANTIC GENERALIZATION OF CONDITIONED AUTONOMIC RESPONSES[1]

PETER J. LANG, JAMES GEER, AND MICHAEL HNATIOW

University of Pittsburgh

Ss were presented with a series of verbal stimuli, some of which were followed by electric shock. Subsequently they were administered unreinforced verbal stimuli, distributed along a scaled meaning dimension extending from words similar to those shocked to neutral words. Responses in 3 autonomic systems were monitored (finger plethysmograph, heart rate, and GSR). Generalized responses were found to words very similar in meaning to the training stimuli. Response amplitude fell off sharply with the presentation of less closely related words.

The conditioning of autonomic responses to a particular "sense or meaning" was first accomplished in Russian laboratories in the late '20s, and the exploration of this phenomenon has continued to be a preoccupation of researchers in that country (Razran, 1961). Salivation, vasoconstriction of the forehead and finger, and possibly blood coagulation have been brought under the control of verbal stimuli. Evidence that these responses then generalized to stimuli of similar meaning has also been obtained.

Work by American experimenters has been more limited, but of no less interest. Razran studied the semantic generalization of the salivation response in 1939. His findings were confirmed by Reiss (1946) in an investigation of the galvanic skin reflex. Lacey and his associates (Lacey & Smith, 1954; Lacey, Smith, & Green, 1955) reported generalization of the heart rate response to rural words, following the pairing of the word "cow" and electric shock. More recently Chatterjee and Eriksen (1960) attempted a re-evaluation of the role of awareness in the generalization of meaningful stimuli.

The current experiment was designed to determine if semantic *gradients* of generalization could be obtained, i.e., if the strength of semantically generalized responses vary as an orderly function of the difference in meaning between test stimuli and stimuli remotely situated on the relevant dimension. To accomplish this *S*s were first presented with a series of verbal stimuli, some of which were followed by electric shock. Subsequently *S*s were administered unreinforced verbal stimuli, distributed along a scaled meaning dimension extending from words similar to those previously shocked to neutral words. Unlike most previous studies, the investigation was not limited to a single dependent variable. The activity in three different somatic response systems (skin resistance, peripheral vascular changes, and heart rate) was measured.

METHOD

Subjects.—The *S*s were 27 undergraduate male volunteers, enrolled in the introductory psychology course at the University of Pittsburgh. Twenty of these students formed the experimental group, and 7 were employed as control *S*s.

---

[1] This research was supported in part by National Institute of Mental Health Grant M-3880.

TABLE 1

Examples of the Four Word Classes Employed as Stimuli and Their Median Hostility Scale Values

High Hostile (HH)		Medium Hostile (MH)		Low Hostile (LH)		Neutral (N)
TORTURE	8.7	CONFLICT	5.1	DISPLEASURE	3.4	NAMELESS
MUTILATING	8.6	SCOLDING	4.8	BLUNT	3.0	RECOMMEND
FURIOUS	7.2	DERISIVE	4.7	CONTEST	1.9	ABSTRACT

Apparatus.—Tone stimuli were initiated by an Eico signal generator. The verbal material was recorded on one track of a Concertone 500 stereo tape recorder. All auditory stimuli were put through a 20-w. General Electric amplifier and delivered to *S* via a high fidelity headset.

The electric shock was provided by a dry cell powered inductorium and administered to the ventral surface of the distal phalanges of the first two fingers of *S*'s left hand. The presentation of the shock was controlled by the second track of the tape recorder. An inaudible electrical signal on the tape started a timing circuit, which presented a 4-sec. shock 4.8 sec. after the appropriate verbal stimulus.

The GSR response was recorded from the second two fingers of the right hand, using the zinc electrodes and zinc sulphate paste recommended by Lykken (1959). The recording instrument was the Fels dermohmeter with write-out on an Esterline Angus milliammeter. The Grass Model 5 polygraph was used to record the standard lead I EKG and the finger plethysmograph. The latter measure was taken from the first finger of the right hand utilizing the method reported by Simeone, Cranley, Grass, Linton, and Lynn (1952).

Verbal stimuli.—The verbal stimuli used in this study were the words scaled for hostility by Buss (1961). College *S*s rated words from 1 (least hostile) to 9 (highest intensity). Words showing excessive variability were eliminated, leaving a final list of 146. In the present study these words were divided into three groups: High Hostile (HH), with a scale value between 8.7 and 6.5; Medium Hostile (MH), scale value between 5.3 and 4.6; Low Hostile (LH), scale value between 3.6 and 1.8. These were supplemented by a pool of Neutral (N) words, judged to be free of hostile connotation by the *E*s. Examples of each word class are presented in Table 1.

Four tape recordings were made by a male speech therapist, who was unaware of the purpose of the experiment.[2] Each consisted of preliminary instructions, a 22-word conditioning series, 3 neutral buffer words, and the 16-word generalization series which provided the basic data of the experiment. The interval between words was 30 sec. No word appeared twice in the same tape. Each conditioning series included 10 neutral words randomly distributed among 12 high hostile words. During conditioning all 12 HH words were followed by electric shock. The 16 words of the generalization series were organized in blocks of 4 that included one member of each word class (HH, MH, LH, N). Order of presentation varied between blocks and was counterbalanced across tapes, so that no word class enjoyed a position advantage in the experiment. Word classes were equated for frequency in the language and syllable length. No shock was administered during the generalization series.

Procedure.—The *S* was seated in a comfortable lounge chair in a partially sound shielded room. After a short period of conversation designed to develop rapport and enlist interest in the experiment, *S* was informed that a "current stimulus" would occasionally be applied to his hand. The level of shock administered during the experiment was adjusted by each *S* at this time. Following attachment of the shock electrodes, *S* was instructed to turn up the attenuator on the inductorium until the sensation experienced was "clearly uncomfortable." "It should be painful, but of course not intolerable—an experience that you find unpleasant, but one that you can stand for a few seconds." Because the attenuator was turned up slowly (allowing temporary adaptation) and the *E* constantly encouraged increases, the levels set by these *S*s were uniformly high. The shock level was generally greater than that set by the *E*s, trying to follow the same instructions.

---

[2] These recordings were made with the help and advice of Aubrey Epstein of the Department of Speech, University of Pittsburgh.

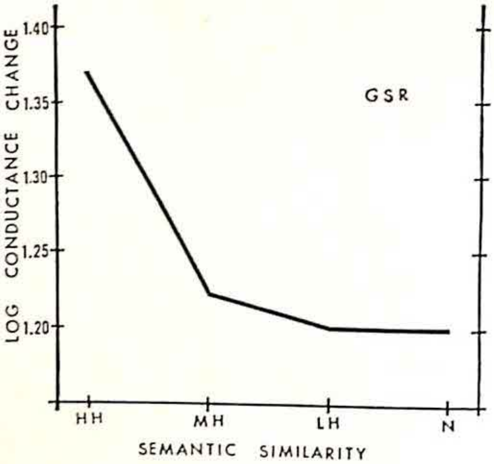


FIG. 1. The mean GSR scores (log conductance change) of the experimental group for the four categories of hostile word stimuli.

After this shock adjustment procedure, the remaining electrodes were attached and the headset placed in position. The *S* was told to sit quietly during the experiment. He was informed that after a short rest period he would hear some tones and then words over the ear phones, and from time to time he would experience the current stimulus to the hand. The *S* was instructed to concentrate on each word and its meaning. He was told that he would be asked to recall the words at the end of the experiment. The *E* then left the subject room and turned off the lights. The *E* monitored *S*'s behavior from an adjacent equipment room which contained the tape recorder and bioelectric recording apparatus.

Following a 1-min. rest, three 800-cps tones at about 85 db. were presented so that *E* could adjust the sensitivity level of the recording apparatus. The *S* subsequently rested for 9 min. after which one of the four prerecorded tapes was presented. The specific tapes were distributed randomly across *S*s with the restriction that each tape was presented to 5 different *S*s. Following the conditioning portion of the tape *E* entered the room, inspected the seating of the electrodes, and told *S* that the experiment would continue for a few more minutes. Three minutes after *E* left the room the buffer words were presented followed by the generalization series.

Control S*s*.—These *S*s were administered only the generalization series and received no electric shock. The same autonomic responses were recorded as in the experimental group.

RESULTS

GSR.—The *S*'s responses to the words of the generalization series were determined in the following manner. The minimum resistance in ohms achieved during the 20 sec. after stimulus onset was subtacted from *S*'s skin resistance when the word was presented. This value was employed as the response measure, unless no decrease in resistance was apparent. In that case *S* was assigned a score of zero. The primary estimate of generalization was based on each *S*'s mean response to all presentations of each of the four word categories. The group means of these data yielded a gradient which extended from a maximum decrease in resistance at the HH words, to a minimum resistance change at the LH and N stimuli. This characteristic was shared by gradients obtained from the group medians (rather than means), and from scores based on resistance changes 10 instead of 20 sec. after stimulus onset. Because the data were not normally distributed, the Friedman variance analysis by ranks was employed to evaluate the significance of this slope. This test yielded a $\chi_r^2 = 10.95$, $p < .02$.

In order to normalize the distribution, and thus justify the use of parametric statistics, the resistance change scores were transformed into units of log conductance change. The semantic generalization gradient described by the transformed data is presented in Fig. 1. The main analysis of variance evaluated the two within-*S* variables, successive blocks and words, and their interaction. It was also designed to analyze three components of the between-words variance: HH vs. (MH + LH + N), N vs. (MH + LH), MH vs. LH. In this way the degree of generalization to members of the reinforced word

TABLE 2

ANALYSIS OF VARIANCE FOR THE GSR RESPONSE

Source	*df*	*MS*	*F*
*S*s	19	4.428	
Blocks (B)	3	3.318	8.07**
B × *S*s	57	.411	
Words (W)	3	.551	2.23
HH vs. (MH + LH + N)	1	1.694	6.70*
N vs. (MH + LH)	1	.005	—
MH vs. LH	1	.008	—
W × *S*s	57	.247	
HH vs. (MH + LH + N) × *S*s	19	.253	
N vs. (MH + LH) × *S*s	19	.228	
MH vs. LH × *S*s	19	.282	
W × B	9	.233	—
HH vs. (MH + LH + N) × B	3	.259	1.15
N vs. (MH + LH) × B	3	.058	—
MH vs. LH × B	3	.334	1.03
W × B × *S*s	171	.246	
HH vs. (MH + LH + N) × B × *S*s	57	.225	
N vs. (MH + LH) × B × *S*s	57	.313	
MH vs. LH × B × *S*s	57	.326	

Note.—Log conductance change scores.
$^{*}p < .05$.
$^{**}p < .01$.

class (HH stimuli) could be compared to the remaining hostile and neutral stimuli, and the most meaningful comparisons among these latter three word classes (MH, LH, and N) would be separately evaluated.[3]

The results of the above analysis are presented in Table 2. The significant *F* associated with blocks indicates the rapid extinction of the GSR response. It will be noted that differences between words are wholly accounted for by the significant difference between the high hostile words and the remaining stimuli taken as a group. There is no evidence of a significant interaction between the two main effects or the component variables.

Heart rate response.—The heart rate response was determined in the following manner. For each *S* the number of interpulse (R-R) intervals between the presentation of the stimulus word and shock onset was determined from the conditioning data. In measuring the responses of the generalization series, an equal number of intervals was counted after the stimulus word, and a 4-sec. period beginning at this ordinal pulse was marked off on the EKG record. The *S*'s response was simply the mean interval (in milliseconds) between all pulses beginning within this 4-sec. period. As others have found (Not-

TABLE 3

MEAN POSTSTIMULUS INTERPULSE INTERVAL AND CARDIAC LABILITY SCORES FOR THE FOUR WORD CLASSES: EXPERIMENTAL GROUP

	HH	MH	LH	N
Interpulse interval	850	836	848	833
Lability score	52	50	50	48

Note.—The interpulse interval is reported in milliseconds. The autonomic lability scores are the same as those described by Lacey and Smith (1954).

---

[3] The authors are indebted to William Bendig, who designed the analysis. We would also like to thank Gary Levinson and Robert Wiater, who assisted with the data reduction and statistical computation.

TABLE 4

Experimental Group Plethysmograph Findings for the Four Word Classes

	HH	MH	LH	N
Mean decrease in finger volume (Blocks I and II)	7.6	4.9	5.6	4.5

Note.—The difference scores are expressed in mm³/5 cm³ of finger volume. $N = 8$.

terman, Schoenfeld, & Bersh, 1952) increasing amplitude of the response is associated with longer interpulse intervals and slower rates; diminished amplitude is represented by shorter intervals and a relatively more rapid rate of cardiac activity.

The primary estimate of generalization is based on each *S*'s mean responses to all presentations of each of the four word categories. The group mean interpulse intervals are presented in Table 3. A Friedman analysis of differences between words was significant well beyond the .01 level ($\chi_r^2 = 13.40$). However, the curve is not monotonic. Although the HH and N words provide the maximum and minimum intervals, respectively, the positions of the MH and LH stimuli are inverted. A second analysis of these data was undertaken utilizing the lability score measure developed by Lacey (1954). Each *S*'s responses are expressed as deviations from a regression line, described by the average pulse during the CS-UCS interval and the average interpulse interval 7 sec. prior to stimulus onset.[4] The slope (Table 3) is in the predicted direction. However, only the HH vs. (MH + LH + N) component of the between-words variance approached significance, $F(1, 19) = 3.08$, $p < .10$.

Finger plethysmograph.—Because of recording problems, the plethysmograph records of 9 *S*s were lost. In order to preserve counterbalancing it was necessary to further reduce the group to 8 (on the basis of a random selection from among the eligible *S*s) before analysis of the data could be undertaken. The response measured was the maximum reduction in finger volume attained during the first wave of vasoconstriction following the presentation of the stimulus (recorded as mm³/5 cm³ of finger volume). As may be seen in Table 4 these findings parallel those obtained from the other two dependent variables. Greater reduction in finger volume tends to be associated with an increase in the semantic similarity of the stimulus to the training words. The means of the first two blocks of words are reported because of the rapid flattening of the gradient which occurred after the first four words were presented. Because of the small number of *S*s with scorable plethysmograph records, none of these data achieved statistical significance.

TABLE 5

Mean Control Group Scores for the Four Word Classes

	HH	MH	LH	N
GSR				
Log conductance change	1.07	1.24	1.30	1.02
Cardiac rate				
Interpulse interval	722	732	727	724
Lability score	51	48	51	49

Note.—All control *S*s are included in all calculations. The plethysmograph data are not presented because, as with the experimental *S*s, recording problems caused the loss of many responses and some entire records. The *S*s for whom complete or partial records were available were too few to meaningfully average. However, a careful inspection of these tracings did not reveal any relationship between response amplitude and word class.

---

[4] The prestimulus measure was taken at this point because, for some *S*s, a procedural artifact compromised the validity of cardiac data immediately prior to the presentation of the CS.

Control Ss.—Inspection of the responses presented by the Control *S*s revealed no gradient or gradient trend in any of the three dependent variables considered in this experiment. Their average scores for heart rate and GSR are presented in Table 5.

DISCUSSION

The data provide clear evidence of the semantic generalization of the GSR response. The *S*s who had received electric shock following the presentation of high hostile words, displayed greater skin resistance changes to other members of this word class (unaccompanied by shock) than to less hostile or neutral words. However, the generalization gradient defined by this response (Fig. 1) is extremely steep. A careful inspection of the raw data reveals little evidence of a differentiating response below a scale value of 5 (the median stimulus of the MH group). This finding is inconsistant with studies of the generalization of instrumental responses to the same words (Buss, 1961; Geer & Buss, 1962). These authors report generalized responses along the entire length of the scale, following a number of different training procedures. While this discrepancy may yet reflect differences in procedure, it also suggests that these response systems differ markedly in the character of their discriminations between stimuli.

Razran (1961) reports the only study of classically conditioned semantic generalization, in which graded variations in similarity between test stimuli and a training word are evaluated. The Russian investigators, Vinogradova and Eysler, employed electric shock as a UCS and recorded changes in the finger and forehead plethysmograph. They found two different responses, depending on the semantic proximity of the test stimulus to the training word. Thus, a highly similar word produced vasoconstriction of both the forehead and digit (the same response observed in the present experiment), whereas words less similar in meaning yielded vasodilation in the forehead plethysmograph. Similarly, the GSR response (which can meaningfully change in only one direction) may be specific to closely related words. Organ systems other than those evaluated here may parallel instrumental behavior, yielding responses to more remote stimuli. The present experiment provides evidence that the heart rate response discriminates between hostile and neutral words. However, different methods of measurement do not yield consistant gradient shapes. It is interesting to note that the response in this study was cardiac deceleration, whereas Lacey and Smith (1954) report the semantic conditioning of cardiac acceleration. This is apparently attributable to method. It will be recalled that Lacey and Smith's *S*s tapped a telegraph key for 15 sec. following the presentation of the training stimuli and then received an electric shock.[5] The procedure used here, and the direction of rate change, is more similar to that of Notterman, Schoenfeld, and Bersh (1952).

The Lacey and Smith findings also suggest that the heart rate response yields better discrimination between related and unrelated stimuli than was obtained in this experiment. However, differences in experimental procedure may very well account for our failure to replicate. In the Lacey and Smith study, *S*s were instructed to chain associate to the training stimulus during the tapping period. This meant that shock followed both the *E*'s CS and a string of related, verbal responses emitted by *S*. It is reasonable to assume that *S*'s associations during the 15-sec. CS-UCS interval overlapped the related words used in the generalization series. If this is true, the experiment could be testing for generalization, using words that had in effect been followed by shock.[6]

---

[5] Davis, Lundervold, and Miller (1957) report that cardiac acceleration is elicited by paced tapping.

[6] John I. Lacy in a personal communication, September 17, 1962 has anticipated this interpretation and evaluated its importance in a subsequent analysis of the original data:

> The chain associations of the subjects were tape recorded. All the words which

> the subject used in association to the "conditioned stimulus" (let us say "cow," in one of our groups) were tabulated. Then, we went through all the words produced in association to the so-called "generalization" stimuli. We tabulated the number of times the subject mentioned words in association to the generalization stimuli that he had previously mentioned in association to the conditioned stimulus. A percentage measure was then taken expressing the communality of associations to the generalization stimuli and to the conditioned stimuli. Thus, an individual with a high percentage of communality tended, in associating to the generalization stimuli, to use the same words he had used for the conditioned stimulus. We then split the group into high and low communality of associations, using a median split. The results were as you might have expected. Those with high communality associations showed beautiful "generalization curves." Those with low communality of association showed poor results. The trend-differences between the high and low communality of the groups were significantly different.

These findings are extremely important, and tempt one to speculate that semantic generalization only occurs when the generalization stimuli are explicitly or implicitly verbalized during conditioning.

A final word should be said about semantic conditioning. It has been frequently stressed that the expression of verbal material is altered if the words have strong autonomic correlates. Interest has been focused on semantic conditioning as an explanation of disturbances of language, e.g., "complex indicators." However, the most important feature of this process is that language can control autonomic activity. In the present instance, a verbal concept (hostility) provided the stimulus for characteristic changes in a number of somatic systems. Recently, Shearn (1961) has shown that human *S*s, attending to their own pulse beat, can learn to avoid electric shock by reducing their cardiac rate. Similarly, the Russian investigator, Lisina (Razran, 1961), reports that *S*s watching their own plethysmograph records, transformed digital vasoconstriction into vasodilation in order to avoid electric shock. In these cases, *S*'s self-instructions (based on learned relations between cardiovascular activity and implicit verbal responses?) initiated important defensive responses. Clearly, the nature of interactions between verbal and autonomic systems deserves the increasing attention of researchers.

REFERENCES

Buss, A. H. Stimulus generalization and aggressive verbal stimuli. *J. exp. Psychol.*, 1961, **61**, 469–473.

Chatterjee, B. B., & Eriksen, C. W. Conditioning and generalization of GSR as a function of awareness. *J. abnorm. soc. Psychol.*, 1960, **60**, 396–403.

Davis, R. C., Lundervold, A., & Miller, J. D. The pattern of somatic responses during a repetitive motor task and its modification by visual stimuli. *J. comp. physiol. Psychol.*, 1957, **50**, 53–60.

Geer, J. H., & Buss, A. H. Supplementary report: Generalization of a nonverbal response to aggressive verbal stimuli. *J. exp. Psychol.*, 1962, **63**, 413–414.

Lacey, J. I., & Smith, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

Lacey, J. I., Smith, R. L., & Green, A. The use of conditioned autonomic responses in the study of anxiety. *Psychosom. Med.*, 1955, **17**, 208–217.

Lykken, D. T. Properties of electrodes used in electrodermal measurement. *J. comp. physiol. Psychol.*, 1959, **52**, 629–634.

Notterman, J. M., Schoenfeld, W. N., & Bersh, P. J. Partial reinforcement and conditioned heart rate responses in human subjects. *Science*, 1952, **115**, 77–79.

Razran, G. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.

Razran, G. The observable unconscious and inferable conscious in current Soviet psychophysiology. *Psychol. Rev.*, 1961, **68**, 81–147.

Reiss, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, **36**, 143–152.

Shearn, R. W. Operant conditioning of heart rate. Unpublished doctoral dissertation, Indiana University, 1961.

Simeone, F. A., Cranley, J. J., Grass, A. M., Linton, R. R., & Lynn, R. B. An oscillographic plethysmograph using a new type of transducer. *Science*, 1952, **116**, 355–357.

(Received July 20, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 559–563

DIFFERENTIAL GSR CONDITIONING AS A FUNCTION OF THE CS-UCS INTERVAL

H. D. KIMMEL AND H. S. PENNYPACKER[1]

University of Florida

Differential GSR conditioning at 4 CS-UCS intervals, $\frac{1}{4}$, $\frac{1}{2}$, 1, and 2 sec., was studied, with 16 Ss in each condition. 16 reinforcements of CS+ and nonreinforcements of CS− were given, the CSs being 40-db. tones of 800 or 1000 cps. The UCS was an electric shock. Nonreinforced trials on CS+ were given after 4, 8, and 16 reinforcements. Differential conditioning increased as a function of the CS-UCS interval. The advantage of the 2-sec. condition was due in part to the low response to CS−. Later in conditioning the GSR to CS+ reduced in the 2-sec. condition but not in the others. These results were attributed to the more rapid growth of inhibition in the longer CS-UCS intervals.

The function relating the strength of simple classical conditioning to the temporal interval between the onsets of the conditioned and unconditioned stimuli (the CS-UCS interval) has been a frequent subject of study (Kimble, 1961). The typical finding of these studies has been that the CS-UCS interval function rises to a maximum in the neighborhood of .5 sec. and falls sharply beyond intervals of 1 sec.

Only recently has any work been reported on the CS-UCS interval in differential conditioning (Hartman & Grant, 1962). In that study CS-UCS intervals up to 1 sec. were used and the degree of differential eyelid conditioning was studied. It was found that the amount of differential eyelid conditioning increased as the CS-UCS interval increased. Hartman and Grant attributed this effect to reduced generalization to the negative CS at the longer intervals rather than to increased responding to the positive CS at these intervals. Their positive CS, in fact, showed the conventional peak in response frequency at about .6 sec.

In the present study four groups of *S*s received differential reinforcement of two auditory CSs of different frequencies, each group having a different CS-UCS interval (and CS duration) during training. Differential conditioning was evaluated by comparing responses on nonreinforced test trials with the positive CS with responses on corresponding trials with the negative CS. It was expected that the groups trained with the longer CS-UCS intervals would require fewer training trials to reach maximal strength of differential conditioning but would also be influenced most later in the conditioning process by inhibitory factors that tend to overcome CR strength of the positive CS. This expectation was based on earlier findings (Grings, Lockhart, & Dameron, 1963; Kimmel, 1959; Meryman, 1953) which, taken together, suggested that the number of trials coinciding with the highest point in the curve of conditioned GSR amplitude is inversely related to the CS-UCS interval in simple classical conditioning.

---

[1] This study was done while the senior author was a Visiting Assistant Professor and the junior author a United States Public Health Service Predoctoral Research Fellow at Duke University.

METHOD

Subjects.—Sixty-four introductory psychology students at Duke University volunteered to serve as *S*s to meet a class requirement. They were assigned randomly to four equal sized CS-UCS interval groups.

Instructions.—The *S*s were informed that two tones differing in frequency and an electric shock would be used. The instructions also stated that the shock would be associated with one of the two tones and that they were to attend to the stimuli to determine with which tone the shock was associated. Previous research (Grings & Kimmel, 1959) and preliminary work in this laboratory suggested that if the *S*s were not instructed in this way, the degree of differential responding to the positive and negative CSs would be reduced. The *S*s were, otherwise, led to believe that this was a study of the effect of environmental stimuli on the GSR.

Apparatus.—The CSs were both 40-db. (re: .0002 dynes/cm^2, rated at the earphone) pure tones produced by General Radio Company equipment and delivered to *S* via Trimm ANB-1 headphones with MX-41-AR cushions. One tone had a frequency of 1,000 cps and the other 800 cps. The determination of which tone was to be the positive CS and which the negative CS was randomly balanced within CS-UCS interval groups.

The UCS was an electric shock delivered via silver electrodes to the fingertips of *S*'s right hand and produced by a Psychological Instruments Company stimulator. Its intensity, for an *S* with a hypothetical resistance of 25,000 ohms, was 50 v. and 2 ma.

Temporal relations of the stimuli were controlled electronically. Each of the four groups had a CS of different duration. These were: 0.35, .60, 1.10, and 2.10 sec., for the positive and negative CSs. The UCS duration was .10 sec. sharing the last .10 sec. of the positive CS's presentation and terminating with it. The four CS-UCS intervals, thus, were: .25, .50, 1, and 2 sec. The CS duration was confounded with the CS-UCS interval intentionally in preference to using a constant CS duration and, consequently, confounding CS-UCS overlap with the CS-UCS interval.

The GSR was picked up by means of an NaCl-zinc-skin junction from the palm and back of the left hand, amplified by a Hunter GSR apparatus, and recorded on an Esterline-Angus ink-writing milliammeter with a paper speed of 3 in/min. The response, defined as the maximum recorder deflection within 4 sec. of the CS onset, was transformed into units of log $(\Delta C + 1)$ for analysis.

Procedure.—Data were collected in a quiet, dark room. The hum of an electric fan masked extraneous sounds. The *E* and the equipment were in an adjoining room. When the *S* arrived at the laboratory, the equipment was attached and the typewritten instructions read by him. Then the door to the experimental room was closed and data collection was begun.

Each *S* first received 2 presentations each of the positive and negative CSs, in ABBA or BAAB order. Then three shocks were delivered, in intensity increments up to the value used during training. Two more presentations were then given of the positive and negative CSs, again in ABBA or BAAB order. The differential conditioning procedure was then begun. This consisted of 16 presentations of the positive CS plus shock and 16 presentations of the negative CS, broken into blocks of eight trials containing 4 presentations of each CS, in ABBABAAB or BAABABBA order. Nonreinforced (test) trials on the positive CS and the negative CS were given following 4, 8, and 16 reinforcements, i.e., following each of the first two eight-trial blocks and following the last eight-trial block. On these extra trials, the order of positive and negative CSs was balanced across *S*s within CS-UCS interval groups.

RESULTS

Figure 1 shows the mean amplitude of the GSR to the positive and negative CSs as a function of the CS-UCS interval, at three stages of training and averaged across all stages. The response values for the positive CS were obtained on the test trials following 4, 8, and 16 reinforcements. The corresponding response values for the negative CS were obtained by averaging the GSRs on negative CS trials No. 4–5–6, 9–10–11, and 17–18–19.

Figure 1 indicates that the amount of differential conditioning was positively related to the CS-UCS interval at all three stages of training and overall, but that the effect became less marked on the third test (following 16 reinforcements). A three-way

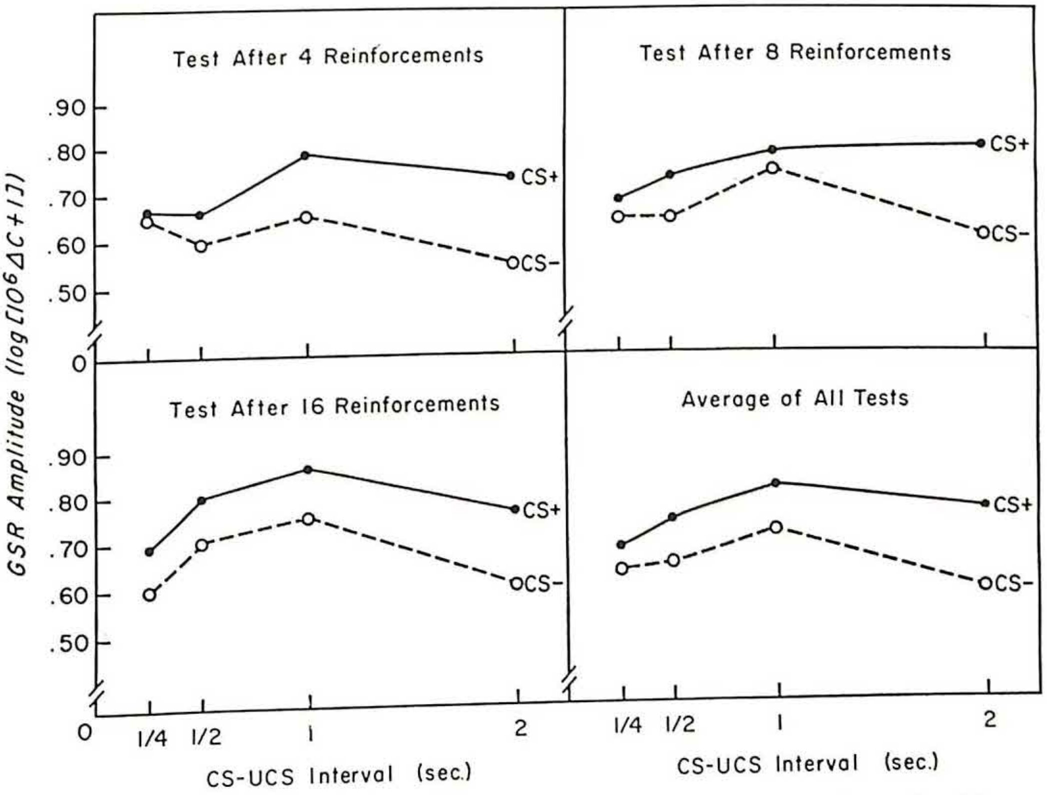


FIG. 1. Mean GSR amplitude (log [$10^6\Delta C + 1$]) to the positive and negative CSs following 4, 8, and 16 reinforcements as a function of the CS-UCS interval.

analysis of variance (Differential Reinforcement, CS-UCS Intervals, and Stages of Training) showed that the effects of Differential Reinforcement and Stages of Reinforcement were both highly significant ($p < .01$) and the interaction of CS-UCS Intervals and Differential Reinforcement was significant ($p < .05$). This analysis is summarized in Table 1.

The degree to which the amplitude of the (generalized) GSR to the negative CS changed over trials during differential reinforcement is shown in Fig. 2. In the 2-sec. condition, the amplitude of the GSR to the negative CS appeared to show less of a progressive increase from trial to trial during training than was shown by the other conditions. In the .50- and 1-sec. conditions the amplitude of the GSR to the negative CS appeared to increase throughout training, while in the .25-sec. condition it appeared to rise for about eight trials and then

TABLE 1

ANALYSIS OF VARIANCE OF GSRS TO POSITIVE AND NEGATIVE CSS AFTER 4, 8, AND 16 REINFORCEMENTS

Source	*df*	*MS*	*F*
Ss	63	0.3438	—
CS-UCS intervals (I)	3	0.2060	*ns*
Ss/within I	60	0.3507	—
Differential reinforcement (DR)	1	0.9842	40.67***
DR × I	3	0.0771	3.19*
DR × *Ss*/within I	60	0.0242	—
Reinforcement stages (RS)	2	0.1270	5.52**
RS × I	6	0.0340	*ns*
RS × *Ss*/within I	120	0.0230	—
DR × RS	2	0.0222	*ns*
DR × RS × I	6	0.0083	*ns*
DR × RS × *Ss*/within I	120	0.0158	—

* $p < .05$.
** $p < .01$.
*** $p < .001$

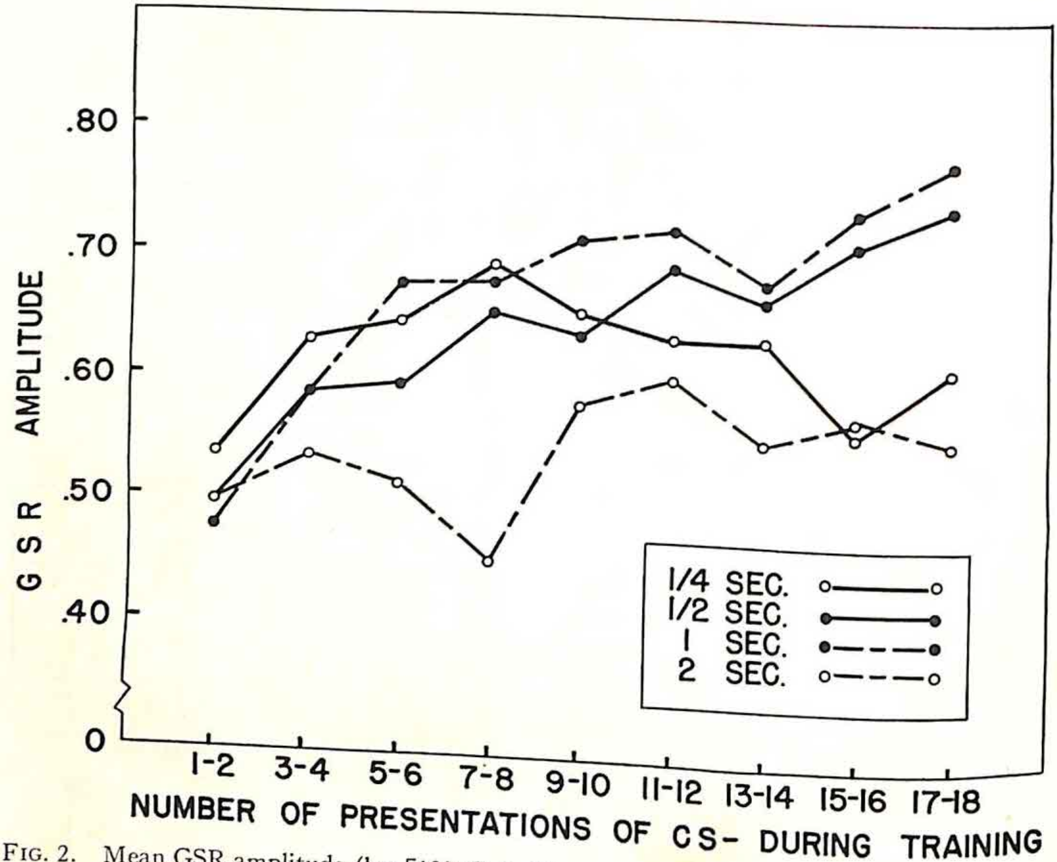


FIG. 2. Mean GSR amplitude (log $[10^6\Delta C + 1]$) to the four negative CSs during differential reinforcement training (averaged in blocks of two trials to increase stability).

fall. Analysis of variance of the GSRs to the negative CS, using only the data from the first and last two-trial stages in Fig. 2, for simplicity, showed that the overall increase in GSR amplitude to the negative CS during training was highly significant, $F\ (1, 60) = 34.39$, and that the interaction of this increase and CS-UCS intervals was also very significant, $F\ (3, 60) = 4.32$, the latter supporting the observation that the change in GSR amplitude to the negative CS differed in the various CS-UCS interval conditions. Error variance for these tests was 0.0277 (in single trial units).

DISCUSSION

Differential conditioning of the GSR was found to increase with increasing CS-UCS intervals. This effect appeared to be due in part to considerably reduced responding to the negative CS in the 2-sec. condition as compared to the other conditions. The superiority of the 2-sec. CS-UCS interval condition tended to be reduced somewhat by further training, however, since it appeared to be least pronounced on the test following 16 reinforcements. This apparent deterioration in differential conditioning in the 2-sec. condition was not manifested in a significant three-way interaction, as would have been required to support a definitive conclusion.

In discussing their more or less similar results in a differential eyelid conditioning experiment, Hartman and Grant (1962) pointed out that the positive CS, considered separately, showed a conventional CS-UCS interval function with a peak near .6 sec. Superior differential conditioning at the longer CS-UCS intervals could, in that study, be attributed to the low level of response to the negative CS at the longer intervals. As Hartman and Grant pointed out, reduced responding to the negative CS at the longer intervals could have been due either to the presence of a more "complete" perceptual response to the positiveness and negativeness of the stimuli or to the possibility that inhibition of a response may require more time. They indicated some preference for the explanation involving an inhibitory concept.

In the data of the present experiment, the conventional CS-UCS interval function for the positive CS was not as readily discernible, although the data seemed to be moving in that direction in the later training trials. Of interest in relation to the possibility that an inhibitory process was responsible for reduced responding to the negative CS in our 2-sec. condition was the fact that only in this condition did the GSR amplitude curve to the positive CS drop between 8 and 16 reinforcements. It is possible that this fall-off in responding to the positive CS in the 2-sec. condition reflected a greater generalization of inhibition from the negative to the positive CS than occurred in the other conditions. It has previously been found (Kimmel, 1959) that a peak in the GSR amplitude curve over reinforced trials occurs after about 6 reinforcements when the CS-UCS interval is 4 sec., probably as a result of the development of inhibition during reinforcement. Meryman (1953), using a trace paradigm and an interval of 19 sec., found a peak after only two or three trials. These earlier findings support the notion that inhibition, at least that which attenuates the GSR to a positive CS during reinforcement, builds up more rapidly over trials when long CS-UCS intervals are used.

REFERENCES

GRINGS, W. W., & KIMMEL, H. D. Compound stimulus transfer for different sense modalities. *Psychol. Rep.*, 1959, **5**, 253–260.

GRINGS, W. W., LOCKHART, R. A., & DAMERON, L. E. Conditioning autonomic responses of mentally subnormal individuals. *Psychol. Monogr.*, 1963, **76**(39, Whole No. 558).

HARTMAN, T. F., & GRANT, D. A. Differential eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1962, **64**, 131–136.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KIMMEL, H. D. Amount of conditioning and intensity of the conditioned stimulus. *J. exp. Psychol.*, 1959, **58**, 283–288.

MERYMAN, J. J. The magnitude of an unconditioned GSR as a function of fear conditioned at a long CS-UCS interval. Unpublished doctoral dissertation, State University of Iowa, 1953.

(Received July 23, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 564–567

EFFECT OF MOTIVATION AND PROGRESS ON THE ESTIMATION OF LONGER TIME INTERVALS[1]

ROBERT D. MEADE

Trinity College

Earlier experiments on short time periods have shown time estimates to be inversely related to rate of progress through a task when there is motivation to reach a goal. This study tested the effects of these variables for 15-, 30-, 45-, and 60-min. periods. A factorial design with 2 rates of progress and 2 levels of motivation was used with 10 Ss under each of the 4 conditions. The experimental task was a stylus maze. The conclusion is that Ss' time estimates are inversely related to rate of progress under high motivation but unrelated under low motivation for all periods except 45 min., where the results fell just short of significance at the .05 level of confidence.

Nearly all experimental investigations of time estimation have dealt with short lengths of time and only a few experiments have studied this process for periods of more than 5 or 10 min. Loehlin (1956) used periods up to 20 min. Sterzinger (1935) reports the effects of drugs on time estimates between 25 min. and 1 hr. Faville (1937), also studying the effects of drugs used a 45-min. period while Bakan (1955) used 60-min. periods. Berman (1939) studied the effect of satiation on time estimates and his *S*s took various amounts of time up to 122 min. However, there has as yet been no systematic study of the perception of longer time intervals and further study is needed.

Recently, studies have shown that certain motivational factors affect the perception of time intervals. Hindle (1951) has demonstrated that perceived time is inversely related to perceived progress rate through a task. Meade (1959, 1960) has shown that when *S*s are motivated to reach a goal, their time estimates are inversely related to perceived progress but that this effect does not hold when *S*s are not motivated to reach a goal.

The present study was designed to determine whether the motivational and rate of progress conditions which affect estimates of short times extend to longer periods up to 1 hr.

METHOD

Design.—The experiments employed a factorial design with two independent variables, motivation and rate of progress, which assumed two values each. There was an experimental replication for four different periods of time; 15, 30, 45, and 60 min.

Subjects.—One hundred sixty male undergraduate college students were used as *S*s in this experiment. For each replication, there were 10 *S*s in each of the four experimental conditions and each *S* was used in one condition only.

Apparatus.—The apparatus consisted of a paper-and-pencil achievement test, a stylus maze, and a blindfold. Timing was done with a stop watch.

Control of Independent Variables

Motivation.—It was intended that the motivation variable be controlled by using two values of ego orientation as described by Iverson and Reuder (1956). By using instructions informing *S*s that the outcome of their work on the maze part of the experiment was a reflection of their intelligence and general ability, it was intended to bring about a condition of high motivation to do well.

---

[1] This investigation was supported by a research grant, M-3193, from the National Institute of Mental Health, Public Health Service.

By using instructions informing *S*s that their work on the maze was of no importance and probably a waste of their effort, it was intended to bring about a condition of low motivation.

The motivation instructions were given to *S* just before he began work on the maze, which was the part of the experiment for which he was to give a time estimate.

Rate of progress.—Prior to starting work on the stylus maze, all *S*s were told that the maze was made up of 40 continuous but irregular regions with the goal at the fortieth region. They were told that they would be given a series of trials to find their way through the maze to the goal region. The *S*s were informed that after each trial, they would be told which region of the maze they had reached to give them some idea of how well they were progressing. In addition, *S*s were told that they would begin working on the maze at the point they reached in a free trial in which they would bring the stylus down on a point in the maze while they were blindfolded.

The *S*s in the slow progress group were told they had reached Region No. 17 on the free trial in the maze. In the 12 trials which followed, they were given scores which indicated they had advanced one region per trial or no region at all. On the twelfth and final trial, they were told that they had reached Region No. 20. On five of the trials, *S*s were told that they were progressing slowly in addition to the contrived scores which *E* gave them.

The *S*s in the fast progress group were told that they had reached Region No. 1 on the free trial in the maze. In the 12 trials which followed, they were given scores which indicated they had advanced one or more regions on every trial. In the latter trials of the series, their advancement through the maze was made to appear to be accelerating by giving scores which indicated advancement considerably above their previous rate. On the twelfth and final trial, they were told that they had reached Region No. 20 which was the same score made by the low progress group on the last trial. On 5 of the trials, *S*s were told that they were progressing rapidly in addition to the contrived scores they were given.

Procedure.—As *S*s were brought into the laboratory, *E*, who was not wearing a watch, remarked that his watch was out of order and asked to borrow *S*'s if he had one. This deprived *S* of an objective means of estimating time. The *S* was then given instructions for the achievement test on which he worked for a period of time equal to that which he would spend on the maze. The achievement test was used to provide *S* with a standard time interval with which he could compare time on the maze portion of the procedure which followed.

When *S* had finished working the required time on the achievement test, he was given the appropriate instructions for the experimental condition under which he was to serve. He was blindfolded and given the trials on the maze in accordance with the progress group of which he was a member. All *S*s had 12 trials, each of which was followed by a rest pause. During the rest period, *E* gave *S* his score for that trial and other information as indicated above.

The lengths of trials and rest pauses for the four different time periods were proportional. For the 15-min. period, trials were 65 sec. in length and rest pauses 10 sec. For 30-min. periods, trials were 130 sec. and rest pauses 20 sec. For 45-min. periods, trials were 195 sec. and rest pauses 30 sec. For 60-min. periods, trials were 260 sec. and rest pauses 40 sec.

At the end of the 12 trials, *S*s were given the following interview:

> 1. Now, you have been working for some time on the paper-and-pencil test and on this maze. Tell me, on which one, the paper test or the maze, does it seem that you were working the longer? I am not interested in your accuracy of making this judgment but only which one actually *seemed* to be *subjectively* longer to you.
>
> 2. All right, now let us say that the paper-and-pencil test was — min. long altogether. You said that the maze seemed to be longer (shorter). Now, using that — min. period as a kind of standard, give me your best estimate in min. and sec. of how much longer (shorter) the maze seemed to you. Again, I am not interested at all in your accuracy but just how much longer (shorter) it seemed to you.
>
> 3. Now, let us consider your goal progress through the maze. What are your feelings about moving in the direction of the goal? Do you think you were making progress in the maze?

Only *S*s who indicated that they felt they were making steady progress toward the maze were retained in the fast progress data and only *S*s who felt they were doing poorly were retained in the slow progress group. The *S*s' responses to Question No. 2 added to (or subtracted from) the amount of time

TABLE 1

Time Estimates of Four Periods as Affected by Motivation and Rate of Progress

Motivation	15 Min.		30 Min.		45 Min.		60 Min.	
	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
Low	16.9	17.2	34.8	33.7	50.8	49.8	64.0	65.5
High	16.1	12.6	31.0	21.7	50.6	45.9	69.0	59.5

taken for the paper-and-pencil test constituted the dependent variable.

RESULTS

The average times for the four conditions of each of the 15-, 30-, 45-, and 60-min. time periods are found in Table 1 and the analysis of variance for these data in Table 2.

The results for the several time periods show that both motivation, rate of progress and the interaction between these two variables produced significant variation in the time estimates. The one exception was for the 45-min. period where the progress variable was significant while both the motivation variable and its interaction with rate fell just short of significance at the .05 level of confidence. The pattern for the 45-min. period follows the same trend as the other time periods, however.

TABLE 2

Analysis of Variance of Time Estimates for Four Periods

Periods	Variable	*df*	*MS*	*F*
15 min.	Motivation (M)	1	201.78	8.25**
	Rate (R)	1	129.38	5.29*
	M × R	1	116.00	4.74*
	Within	36	24.45	
	Total	39	35.19	
30 min.	M	1	524.00	8.08**
	R	1	383.00	5.90*
	M × R	1	326.10	5.03*
	Within	36	64.87	
	Total	39	91.50	
45 min.	M	1	429.30	4.06
	R	1	437.30	4.13*
	M × R	1	432.20	4.09
	Within	36	105.76	
	Total	39	130.92	
60 min.	M	1	984.00	4.38*
	R	1	948.00	4.22*
	M × R	1	1,051.00	4.68*
	Within	36	224.64	
	Total	39	283.85	

$*p < .05$.
$**p < .01$.

DISCUSSION

Inspection of the data shows that they are basically in agreement with the results obtained for shorter periods (Meade, 1959). The time estimates for the low motivation groups show no apparent effect of the rate of progress variables. In two cases, the 30- and 45-min. periods, time estimates under rapid progress were slightly shorter than those under slow progress while in the other two cases, the 15- and 60-min. periods, the reverse was true. This would be expected in random sampling if the variable of progress were not effective.

On the other hand, the average time estimates made under the condition of high motivation were shorter under rapid progress than they were for slow progress. As in the earlier studies, then, time estimates are inversely related to rate of progress only under conditions where *S* has motivation to do well and reach a goal but unrelated to rate of progress where there is low motivation.

The findings of this study may be considered from a number of different standpoints, two of which will be considered here. One of these concerns the problem of boredom. Subjects working under conditions of low motivation are in exactly the situation which brings on the reaction of boredom or satiation. Even in the rapid progress groups, *S*s are unaffected by the information that they are moving toward a point in the task which appears to be useless to them. They are not concerned with whether they are or are not doing well on a useless task and indeed may actually pay no attention to information relevant to their progress. Regardless of what they do, it may be regarded as accomplishing nothing setting the stage for the appearance of boredom.

However, when *Ss* are motivated and interested in what they are doing, the acquisition of information relevant to their progress is of some concern to them. It is in the condition where they are motivated but seeming to make little or no progress that *Ss* may regard themselves as accomplishing nothing and this is the same situation as in both progress conditions of the low motivation groups. The data show that the slow progress condition of the high motivation level produces time estimates differing little from that of either progress condition of the low motivation level.

Another way of considering these data introduces the allied problem of frustration (defined here as basically the result of goal blockage). Subjects in three of the four experimental conditions can be regarded as working under conditions designed to produce frustration. Both of the progress groups in the low motivation condition and the slow progress group of the high motivation level are accomplishing little or nothing and hence are blocked from reaching a goal. These three conditions, as was pointed out, produced time estimates which varied little. At the same time, those free of the frustrating conditions produced shorter time estimates.

When the results are considered in any of these ways, both motivation and progress toward goals seem to be of importance in making time estimates. Thus far, research has been limited to approach motivation and forward progress toward goals. Since these variables have been found to affect time estimates, further experimentation is needed. Additional experiments are called for in such problems as the effect of avoidance motivation and progress away from goals.

REFERENCES

BAKAN, P. Effect of set and work speed on time estimation. *Percept. mot. Skills*, 1955, **5**, 147–148.

BERMAN, A. The relation of time estimation to satiation. *J. exp. Psychol.*, 1939, **25**, 281–293.

FAVILLE, M. La percezione del tempo nell' embrezza mescalinica. *Rass. Studi psichiat.*, 1937, **26**, 455–462.

HINDLE, H. Time estimates as a function of distance traveled and relative clarity of a goal. *J. Pers.*, 1951, **19**, 483–501.

IVERSON, M., & REUDER, M. Ego-involvement as an experimental variable. *Psychol. Rep.*, 1956, **4**, 147–181.

LOEHLIN, J. The influence of different activities on the perception of time. Unpublished doctoral dissertation, University of California, Berkley, 1956.

MEADE, R. Time estimates as affected by motivational level, goal distance, and rate of progress. *J. exp. Psychol.*, 1959, **58**, 275–279.

MEADE, R. Time estimates as affected by need tension and rate of progress. *J. Psychol.*, 1960, **50**, 173–177.

STERZINGER, O. Chemopsychologische untersuchungen uber den Zeitsinn. *Z. Psychol.*, 1935, **134**, 100–131.

(Received July 25, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 568–571

EFFECTS OF WITHIN-SESSION INCENTIVE CONTRAST ON INSTRUMENTAL ACQUISITION AND PERFORMANCE[1]

W. A. PIEPER[2] AND MELVIN H. MARX
University of Missouri

3 groups of 10 rats each were magazine trained and bar press trained daily over 28 days. 10 min. of 1-min. fixed interval magazine presentations were followed by 10 min. of 1-min. fixed interval bar pressing. The group whose incentive was shifted up from magazine to bar press (4–11.3% sucrose) showed an initial depression of acquisition bar press rate but an eventual asymptotic facilitation, compared with the 11.3/11.3% controls. The group whose incentive was shifted down (32/11.3%) showed an initial (but unreliable) facilitation and an asymptotic depression. Interpretation of these results is suggested in terms of the development of discrimination between magazine and bar press incentives.

In reviewing the incentive-shift literature, Collier and Marx (1959) pointed out that in most previous studies (e.g., Crespi, 1942; Spence, 1956; Zeaman, 1949) the shift in amount of reinforcement occurred after asymptote had presumably been reached, and *S*'s experience with the first magnitude of incentive was concurrent with acquisition of the instrumental response. Such confounding of instrumental-response acquisition with the original incentive magnitude was controlled in the Collier-Marx study by keeping the experience of the first incentive magnitude independent of the acquisition of the instrumental response. This was accomplished by magazine training all *S*s for 8 days using the primary incentive magnitudes of 4%, 11.3%, or 32% sucrose, depending upon the group. All groups then received 10 daily 20-min. BP (bar press) training sessions using the same incentive as reinforcement (11.3% sucrose).

Marx and Pieper (1962) modified the relationship between incentive shift and instrumental behavior by running daily test sessions in which direct feeding on one incentive value was regularly alternated with acquisition of the instrumental response on the second, contrasting incentive value. Reinforcement was continuous. The major procedural advantage gained from this design was the continuing contrast of incentive values over the entire acquisition period. An unexpected facilitation of bar pressing occurred when the incentive shift was from 64% to 32% sucrose. In a further study (Marx & Pieper, 1963), this facilitation with a shift down in incentive was confirmed, using a 1-min. fixed interval schedule, but a subsequent depression of bar pressing rate also occurred.

The present experiment concerns a further modification of these temporal relationships, and the addition of a shift-up group, within the same general procedural framework. Daily test sessions included 10 min. of direct feeding one incentive value followed immediately by 10 min. of bar pressing (BP) for the second, contrasting value. The direct presentations each day utilized 32%, 11.3%, and 4% sucrose solution with all *S*s being shifted to 11.3% for BP training. These sucrose concentrations are the same as those used in the Collier-Marx (1959) study.

---

[1] This research was supported in part by United States Public Health Service Grant MH 00817-09 from the National Institute of Mental Health. We are grateful to James Miller for his careful collection of the data.

[2] United States Public Health Service Predoctoral Research Fellow.

METHOD

Apparatus.—Eight liquid reinforcement test boxes were used. They were equipped with magazines delivering .10 ml. liquid reinforcements. These magazines consisted of 12-in. diameter aluminum plates with 72 $\frac{5}{16}$-in. cups drilled around the periphery of each plate. Each cup was loaded with sucrose solution using a Cornwall automatic pipetting device. A second 12-in. plate constructed of plastic was used as a cover to reduce evaporation. The magazines were advanced by a Geneva movement with a motion which was barely audible. A Gerbrands bar requiring approximately 22 gm. of force was used to activate the magazine, which was located below the rat cage and slightly to one side of the bar. This cage was a Hoeltge HB-11A, 8 × 9¾ × 7 in., of a type similar to that which housed the animal. Each individual assembly was located in a ventilated, sound-proofed, light-tight ice-box hull dimly illuminated inside by a 7½-w. light bulb.

Subjects.—Thirty-six male albino rats of the Sprague-Dawley strain were obtained from the Holtzman Company and placed on food deprivation 10 days prior to beginning the experiment. All *S*s were maintained upon 10 gm. dry Purina laboratory chow mash fed approximately 30 min. after running. Water was freely available at all times. The caloric value of the 10 gm. of laboratory chow plus sugar intake was approximately 38 calories per day per animal. This restricted intake produced a mean weight loss of approximately 20% of *S*'s ad lib. feeding weight. The *S*s were 110 days of age and experimentally naive at the start of the experiment.

Procedure.—Three days of magazine (M) accommodation preceded bar press (BP) training. The 36 *S*s were first randomly assigned to three sucrose-concentration groups (4, 11.3, and 32% by weight in water). The first and subsequent accommodation sessions lasted for a total of 20 min. Each *S* periodically received one .30-ml. sucrose reinforcement per minute of the assigned concentration. The second session repeated the procedure of the first with the exception of a reduction in reinforcement volume to .20 ml. The third day this volume was still further reduced to a final value of .10 ml. This procedure was instituted because the larger solution volumes more nearly filled the $\frac{5}{16}$-in. cups and thus facilitated accommodation to the magazine. A minimum of 10 *S*s per group met the criterion of taking 19 out of the 20 possible reinforcements on the third magazine accommodation day. These 30 *S*s were introduced the next day to BP training, which continued through 28 days of testing.

The BP training procedure combined an intersession shift from magazine training to BP training with a corresponding intersession shift in reinforcement concentration. To accomplish this, M training was reduced to 10 min. with an immediate shift to BP training for 10 min. The M reinforcement concentration remained the same as during accommodation (4, 11.3, or 32%) but all *S*s were shifted to 11.3% reinforcement for BP training using a 1-min. fixed-interval reinforcement schedule. Reinforcement volume was held constant at .10 ml. for both M and BP training.

The shift to BP training was facilitated by using two cages for each test box: one without a bar for M training, and one with the bar installed for BP training. The time required for the shift was held constant at 1 min.

RESULTS

The mean numbers of BPs for 2-day test blocks are shown in Fig. 1. Both the acquisition and performance of the instrumental response appear to be clearly related to the contrasting incentive presented during the 10 min. of M training given each day. An analysis of variance of the data for the 28 test days was performed (within-*S*s df reduced by 12 because of missing data in some cells). This analysis indicated that the main effect of Groups was reliable, F (27, 717) $= 14.70$, $p < .01$, as was the Trials × Groups interaction, F (54, 717) $= 6.52$, $p < .01$.

Individual curves, when examined over trials, demonstrated essentially the same results.

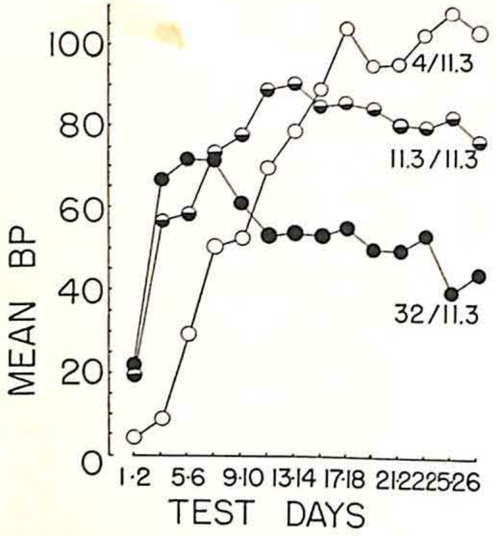


FIG. 1. Mean number of bar presses.

In order to evaluate the early acquisition and the asymptotic performance separately, the first 6 days and the last 6 days were individually analyzed. A simple analysis of variance indicated that the groups were reliably different at both stages—first six trials, $F(2, 27) = 9.05$, $p < .01$; last six trials, $F(2, 27) = 16.67$, $p < .01$. Dunnett's test for two-sided comparisons (Edwards, 1960, pp. 152 f.) was used to evaluate the differences between each of the experimental groups and the control group. The inferiority in acquisition of the 4/11.3 group was reliable, $t(2, 24) = 3.16$, $p < .01$. The superiority of the 32/11.3 group, however, was not reliable. Both asymptotic comparisons were reliable—4/11.3 group, $t(2, 24) = 2.45$, $p < .05$; 32/11.3 group, $t(2, 24) = 3.23$, $p < .01$.

DISCUSSION

In this experiment, direct feeding of a less preferred incentive produced slower instrumental acquisition, but enhancement of acquisition by a directly-fed preferred incentive (Marx & Pieper, 1962, 1963) was not obtained at a statistically satisfactory level. Asymptotic performance of each group clearly demonstrated the eventual development of the "undershooting" and "overshooting" effect, replicating and extending the findings of Marx and Pieper (1963) with the same 1-min. fixed-interval schedule.

The early facilitation or depression of the acquisition of the instrumental response may be interpreted as a result of *S*'s failure to discriminate between the direct-feeding sucrose concentration and the BP concentration. This failure to discriminate results in either enhancement or reduction in effective incentive value of the BP incentive.

The undershooting and overshooting performance results may likewise be attributed to a somewhat similar discriminatory mechanism. The *S*s shifted down respond to the devaluation of the BP incentive by reduced instrumental activity relative to the controls; the *S*s shifted up respond to the more highly preferred incentive with increased instrumental activity when compared with the controls.

A strictly physiological explanation of the results in terms of postingestive or taste factors (McCleary, 1953) is contraindicated by the reversal of the relative positions of the groups as they go from acquisition to asymptotic performance. McCleary, using direct stomach loading of several solutions of varying concentration, found that the tonicity of the load per se was a significant controlling factor regulating the amount of sugar test solution drunk either immediately or after a 15-min. delay. Although the effect of prefeeding with solutions of different tonicity might possibly explain the undershooting and overshooting results at asymptote, the opposite relationships in early acquisition cannot be easily accounted for on the same basis. For this reason also taste factors cannot be used as a comprehensive explanation. It appears that a biphasic explanation, such as the one suggested above, is required.

Although the early facilitation effect in the shift-down group, previously observed by Marx and Pieper (1963), did not reach statistical reliability in the present experiment, the reliably slower acquisition and reliably superior asymptotic performance of the 4/11.3 group provides a strikingly symmetrical contrast with the superior acquisition and inferior asymptotic performance of the 32/11.3 group. This extension of earlier work to include the shift-up condition thus increases the generality of the motivational effects thus far obtained.

REFERENCES

COLLIER, G., & MARX, M. H. Changes in performance as a function of shifts in the magnitude of reinforcement. *J. exp. Psychol.*, 1959, **57**, 305–309.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart, & Winston, 1960.

MARX, M. H., & PIEPER, W. A. Acquisition of instrumental response as a function of incentive contrast. *Psychol. Rep.*, 1962, **10**, 635–638.

MARX, M. H., & PIEPER, W. A. Instrumental acquisition and performance on fixed-interval reinforcement as a function of incentive contrast. *Psychol. Rep.*, 1963, **12**, 255–258.

MCCLEARY, R. A. Taste and postingestion factors in specific-hunger behavior. *J. comp. physiol. Psychol.*, 1953, **46**, 411–421.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received July 25, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 572–575

EFFECT OF SEX OF SUBJECT, SEX OF EXPERIMENTER, AND REINFORCEMENT CONDITION ON SERIAL VERBAL LEARNING[1]

MAVIS HETHERINGTON AND LEONARD E. ROSS
University of Wisconsin

Sex of S, sex of E, and verbal reinforcement condition were investigated in a serial trigram learning situation. 10 male and 10 female Es each ran 24 Ss—12 males and 12 females equally divided into verbal reward, neutral, and verbal punishment conditions. Sex of E and interactions involving this factor were not significant, but the E and E × Reinforcement Condition × Sex of S variables, were significant. A significant Sex of S × Reinforcement Condition interaction was found. Male performance differences across reinforcement conditions were not significant but females learned significantly faster under reward or punishment than under neutral conditions. Females learned significantly more rapidly than males only under the punishment condition.

The variables of sex of *S*, sex of *E*, and verbal reinforcement condition have been inadequately studied in verbal learning. An experiment by Archer, Cejka, and Thompson (1961), using a serial trigram task, found that the performance of male *S*s was not influenced by sex of *E* but that female *S*s learned significantly more rapidly with a male *E* than with a female *E*. However only one female and two male *E*s were used in this study, and specific individual personality differences may have been confounded with sex.

Results of studies of the interaction of sex of *S* and sex of *E* in other learning situations have been inconsistent (Binder, McConnell, & Sjoholm, 1957; Ferguson & Buss, 1960; Gewirtz, 1954; Gewirtz & Baer, 1958; Stevenson, 1961). These inconsistencies may be a result of differences in learning tasks and ages of *S*s utilized, or to an inadequate sampling of *E*s in all except the Stevenson study.

An interaction between sex of *S* and *E* would have important implications for the findings of past verbal learning studies where such factors often have not been controlled or reported. The present study, therefore, was designed to investigate the effect of these variables, under varying conditions of verbal reward and punishment, in a serial trigram learning task.

METHOD

Subjects and experimenters.—The *S*s were 240 male and 240 female students from introductory psychology courses at the University of Wisconsin. Ten male and 10 female psychology majors, enrolled in advanced research courses, served as *E*s.

Design and procedure.—The 480 *S*s were assigned at random to the cells of a 2 × 2 × 3 design, one dimension representing the sex of *S*s, one of the sex of *E*s, and one the reward-neutral-punishment reinforcement conditions. The 10 male and 10 female *E*s each ran 24 *S*s, 12 males and 12 females, with 4 of each sex under each reinforcement condition.

Each *S* learned a list of nine trigrams using the serial anticipation method. The trigrams used, and their association values (Archer, 1960) were as follows: CIS (80%), NYZ (20%), VOD (50%), GIM (80%), WEF (20%), SYX (50%), ROH (50%), KAW (80%), and FIQ (20%). No significant difference between the

---

[1] Supported in part by the Research Committee of the Graduate School from funds supplied by the Wisconsin Alumni Research Foundation.

TABLE 1

Reward and Punishment Statements

Reward	Punishment
You're doing fine.	Let's hope you do better next time.
Good. Let's try again.	You had a lot of errors that time.
That was very good.	Try again.
You're improving quickly.	Are you really trying?
That was a good trial.	You're a slow learner.
You're a fast learner.	Wrong again.
Very good.	No you're still making mistakes.
You're doing better all the time.	You're very slow.
Fine.	Come on now. Concentrate!
Keep up the good work.	Don't you ever learn?
That was a good one.	You're not making much progress.
You've almost got it.	This could go on forever.
That was close.	Do you always have this much trouble with simple problems?
You're doing very well.	We'll be here all day at this rate.
You're one of the fastest subjects I've had today.	You're the slowest subject I've had today.
Excellent!	My God, you students are dumb!
You'll have it in no time.	Wrong.
You're close.	You have no place to go but up.
Good.	Will you please try a little harder?
That's the way to go.	Well here we go again.

sexes was found by Archer for any of the trigrams. The trigrams were presented on a standard memory drum for a 2-sec. exposure, with 2 sec. between syllables and a 2-sec. intertrial interval. Learning was continued to a criterion of one perfect trial.

The *Ss* were initially instructed to spell each syllable before it appeared and to learn the list as rapidly as possible. The reinforcement conditions were defined in terms of the comments made by the *E* to *S* during the intertrial interval. Twenty statements were prepared for the reward and punishment conditions while the *E* remained silent during the intertrial interval under the neutral condition. These statements are listed in Table 1 in the order in which they were presented. If the *S* had not met the learning criterion by 20 trials the list was repeated. The same order of statements was used for all *Ss* in each reinforcement group.

All experimental sessions were recorded and spot monitored to insure that *Es* were following the procedure and using the statements properly.

Results and Discussion

Table 2 presents the mean trials to criterion for each of the 12 sex of *E*, sex of *S*, and reward-combination groups. The summary of an analysis of variance of these data is reported in Table 3.

The finding of no significant effect of sex of *E*, or interactions of this factor with the other variables, is not in agreement with the Archer, Cejka, and Thompson (1961) study. Since the present experiment used a more adequate sample of *Es*, 10 females and 10 males as compared to 1 female and 2 males, the Archer et al. results were probably due to inter-*E* variables other than sex. The significant *E*

TABLE 2

Mean Trials to Criterion as a Function of Sex of *S*, Sex of *E*, and Reinforcement Condition

Sex of *S*	Sex of *E*	Reward Cond.	Neutral Cond.	Punishment Cond.
Female	Female	22.63	27.78	24.13
	Male	24.05	29.90	21.00
Male	Female	24.83	27.10	28.60
	Male	26.70	24.63	29.07

TABLE 3

Analysis of Variance of Trials to Criterion

Source	*df*	*MS*	*F*
Sex *E* (*E*)	1	.30	
Sex *S* (*S*)	1	437.01	5.59*
Reinforcement (R)	2	316.94	2.04
Individual *E* (I)	18	228.00	1.83*
$E \times S$	1	1.01	
$E \times$ R	2	90.03	
$S \times$ R	2	863.64	4.70*
$S \times$ I	18	78.18	
R $\times$ I	36	155.02	1.24
$E \times S \times$ R	2	171.10	
I $\times S \times$ R	36	183.73	1.47*
Within	360	124.78	

* $p < .05$.

effects of the present study support this interpretation, and are in agreement with the results of Stevenson (1961) who found that some *E*s were consistently more reinforcing than others with children in a marble game situation, although they were working under the same verbal reward condition. While Stevenson was unable to relate these *E* differences to age, appearance, experience with children, scholastic status, or scores on the MMPI, more systematic experimental investigations of personality variables might be fruitful in discovering the relevant factors.

Of interest is the finding of a significant interaction between sex of *S* and reinforcement condition. To further analyze this interaction, *t* tests were used to compare the performance of males and females at each reinforcement condition, and simple analyses of variance were performed to compare each sex across reinforcement conditions.

The male and female groups did not differ significantly under reward or neutral reinforcement conditions, $t\ (158) = 1.49$, $p > .10$; and $t\ (158) = 1.92$, $p < .10$, respectively, but females showed superior performance under the punishment condition, $t\ (158) = 4.05$, $p < .01$. Male performance differences across reinforcement condition were not significant, $F\ (2, 237) = 1.59$, $p > .20$, but females did differ as a function of reinforcement condition, $F\ (2, 237) = 8.31$, $p < .001$. A Duncan multiple range test indicated that with females both the reward and punishment groups differed from the neutral group at the .01 level, but that the female reward and punishment groups did not differ significantly.

The results of this study indicate that although sex of *E* is not an important variable in serial verbal learning, other *E* characteristics do affect verbal learning performance. This poses a methodological problem for while it is relatively easy to control and balance the sex of *E*s across experimental conditions, these other *E* characteristics are as yet unidentified and exist as unknown factors in the experimental situation. Until these *E* characteristics are identified a possible, though cumbersome, solution to this problem would be to use a random sample of *E*s rather than a single *E* as is usually the case in verbal learning studies.

The significant Sex of $S \times$ Reinforcement Condition interaction demonstrates the importance of controlling these factors in verbal learning. Since the *E*'s comments and other social interaction variables would be expected to differentially affect the performance of male and female *S*s it seems necessary to strictly specify and control these features of the experimental situation. The data suggest that social interaction, whether involving praise or censure, improves the performance of females whereas the performance of males is largely independent of this factor.

References

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

ARCHER, E. J., CEJKA, J., & THOMPSON, C. P. Serial-trigram learning as a function of differential meaningfulness and sex of subjects and experimenters. *Canad. J. Psychol.*, 1961, **15**, 148–153.

BINDER, A., MCCONNELL, D., & SJOHOLM, N. Verbal conditioning as a function of experimenter characteristics. *J. abnorm. soc. Psychol.*, 1957, **55**, 309–314.

FERGUSON, D. C., & BUSS, A. H. Operant conditioning of hostile verbs in relation to experimenter and subject characteristics. *J. consult. Psychol.*, 1960, **24**, 324–327.

GEWIRTZ, J. L. Three determinants of attention-seeking in young children. *Monogr. Soc. Res. Child Develpm.*, 1954, **19**(2, Whole No. 59).

GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56.

STEVENSON, H. W. Social reinforcement with children as a function of CA, sex of *E*, and sex of *S*. *J. abnorm. soc. Psychol.*, 1961, **63**, 147–154.

(Received July 28, 1962)

ERRATUM

In the article "Wave-Length Generalization as a Function of Spacing of Test Stimuli" by Herbert Friedman (*J. exp. Psychol.*, 1963, **65**, 334–338) the sixth line in the first paragraph on page 334 should read:

units that separate the test stimulus from the training stimulus.

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 576–582

INCREASE OVER TIME IN THE STIMULUS GENERALIZATION OF ACQUIRED FEAR[1]

WALLACE R. McALLISTER AND DOROTHY E. McALLISTER

Syracuse University

Following classical fear conditioning, Ss were trained to jump a hurdle to escape the fear-eliciting stimuli. 2 groups were conditioned in the start box of the hurdle apparatus (Same CB) and 2 in a separate, but similar, conditioning box (Diff CB). Under each condition, 1 group had hurdle training after a 3-min., and the other, after a 24-hr., postconditioning delay. For the 2 Diff CB groups, hurdle training involved a generalized stimulus situation. All groups, except Group 3 min.-Diff CB, learned to jump the hurdle in 25 trials. The significant improvement in performance with an increase in postconditioning delay under the Diff CB, but not under the Same CB, condition was attributed to an increase in the strength of the fear response over time to generalized, but not to the original, stimuli.

In a recent acquired drive study, it was found that immediately following classical conditioning procedures in which a CS was paired with shock, *S*s did not learn a hurdle-jumping response to escape the CS. If, however, a period of 24 hr. intervened between conditioning and hurdle-jumping training, learning took place rapidly (McAllister & McAllister, 1962a). Theoretically, fear as a response should be classically conditioned to the CS in the first phase of the experiment.[2] Escape from the CS in the second phase would, presumably, reduce the fear and thus reinforce the learning of hurdle jumping. Since the theory would predict such learning regardless of the length of the postconditioning delay, an explanation of the effect of this variable was sought. One of the procedures adopted for convenience was to condition *S*s in a box which was a replica of the start box of the hurdle-jumping apparatus. To the extent that these boxes differed, hurdle jumping was conducted in a generalized stimulus situation. It is on this basis that an explanation of the previous results can be made.[3]

Although not studied extensively, an effect of time on the slope of stimulus generalization gradients has been reported. For instance, Perkins and Weyant (1958) found that the running response to a food reward generalized between alleys of different colors to a greater extent after 1 wk. than after 1 min. If their finding applies to fear responses, the amount of fear elicited by the generalized stimuli in the start box of the hurdle apparatus might be inadequate to serve as the basis for learning immediately following conditioning. The increased amount of generalized fear expected after a delay interval might, however, be sufficient. The purpose of the present experiment was to test this interpretation of the effect of the postconditioning delay variable.

---

[1] This investigation was supported in part by a research grant, M-2064, from the National Institute of Mental Health, Public Health Service. The authors wish to thank Ann Cooper and Joseph J. Franchina for aid in the collection and analysis of the data. A report of these results was presented at the 1962 Eastern Psychological Association meetings.

[2] It has been demonstrated that the effective CS in this situation is a compound stimulus since fear is conditioned both to the discrete CS (light) and to the static cues of the conditioning box (McAllister & McAllister, 1962b).

[3] The authors are indebted to C. C. Perkins, Jr. for suggesting this explanation.

METHOD

Subjects and design.—Sixty-four female, naive, hooded rats, 91–129 days of age, from the colony maintained by the Psychology Department of Syracuse University, were used. Two additional *Ss* were discarded because of apparatus failure. The *Ss* were randomly paired and then the pair was assigned at random to one of two postconditioning delays. One *S* of each pair was randomly assigned to be conditioned in the start box of the hurdle-jumping apparatus in which they were later to be trained. The other *S* of the pair was conditioned in a shock box, different from, but similar to, the start box of the hurdle-jumping apparatus in which they were later trained. Thus, there were four groups: two in which the conditioning box was the same as the start box of the hurdle-jumping apparatus (Same CB) and two in which it was different (Diff CB). Under each condition one group was given hurdle-jumping training with a 3-min., and the other with a 24-hr., postconditioning delay.

Apparatus and procedure.—A hurdle-jumping apparatus and a conditioning box were used. The hurdle apparatus consisted of a start box and a safe box, each $9\frac{3}{4}$ in. long × $4\frac{1}{2}$ in. wide × 5 in. high, separated by a $\frac{3}{4}$-in. partition. Set in the partition, resting on a 2-in. high hurdle, was a guillotine door $2\frac{1}{2}$ in. wide × 3 in. high. The start box, painted white, had a grid floor; the gray safe box, a wooden floor, hinged so as to serve as a floor switch. Above each of the boxes was another box with an opal-glass bottom, which contained light sources for the CS and the intertrial illumination and also served as a cover. Although the conditioning box was constructed to be a replica of the start box of the hurdle apparatus, some differences were present. Those discernible to the human observer consisted of the visibility, through the grids, of a gray, wooden support at one end of the start box but not of the conditioning box and some variation in appearance between the guillotine door of the hurdle apparatus and the simulated door in the conditioning box. In addition, since the light sources did not pass through a perfectly diffusing medium, there may have been discriminable differences in the light patterns on the glass, although the illumination measured in each box was approximately the same.

The *Ss* were conditioned in pairs, one in the conditioning box and one in the start box of the hurdle apparatus. The boxes were wired in parallel. The CS, of 6-sec. duration, was an increase in illumination from 7 ft-c (intertrial illumination) to 115 ft-c. The UCS was a 2-sec. shock of 125 v. delivered to the grids successively at a rate of two impulses per grid per sec. through a 100,000-ohm resistance. Presentation of the UCS occurred 4 sec. following the onset of the CS, and both terminated simultaneously.

On each of the first 2 days of the experiment, all *Ss* were handled for 20 min. and were allowed to explore each side of the hurdle apparatus for 10 min. On the third day, 35 conditioning trials were administered with a 2-min. intertrial interval. Following conditioning, *Ss* were returned to their home cage for the postconditioning delay. Twenty-five hurdle-jumping trials were administered following the delay and again on the next day. These days will be referred to as Day 1 and Day 2 throughout the paper. Since *Ss* were run in pairs, the postconditioning delay for the first *S* of a pair was exact; for the second *S* the delay was lengthened by the time required for the first trial of the first *S* (a maximum of 110 sec.).

The hurdle-jumping trials were administered with a minimum intertrial interval of 30 sec. On each trial *S* was placed in the start box and 10 sec. later the CS was presented simultaneously with the raising of the guillotine door. Latency was measured with a Hunter Klockounter, in .01 sec., from the opening of the door to the depression of the floor switch in the safe box. After jumping, *S* remained in the safe box for 10 sec. and was then removed to a separate holding box for the intertrial interval. If no jump occurred within 60 sec., *S* was removed to the holding box and a 60-sec. latency recorded. On Day 2 if *S* did not jump within 60 sec. on 10 consecutive trials, training was terminated, and 60-sec. latencies were recorded for the remaining trials.

Further details of the apparatus and general procedure are reported elsewhere (McAllister & McAllister, 1962a).

RESULTS

In Fig. 1 are presented the means of reciprocals of latency of hurdle jumping in blocks of five trials for

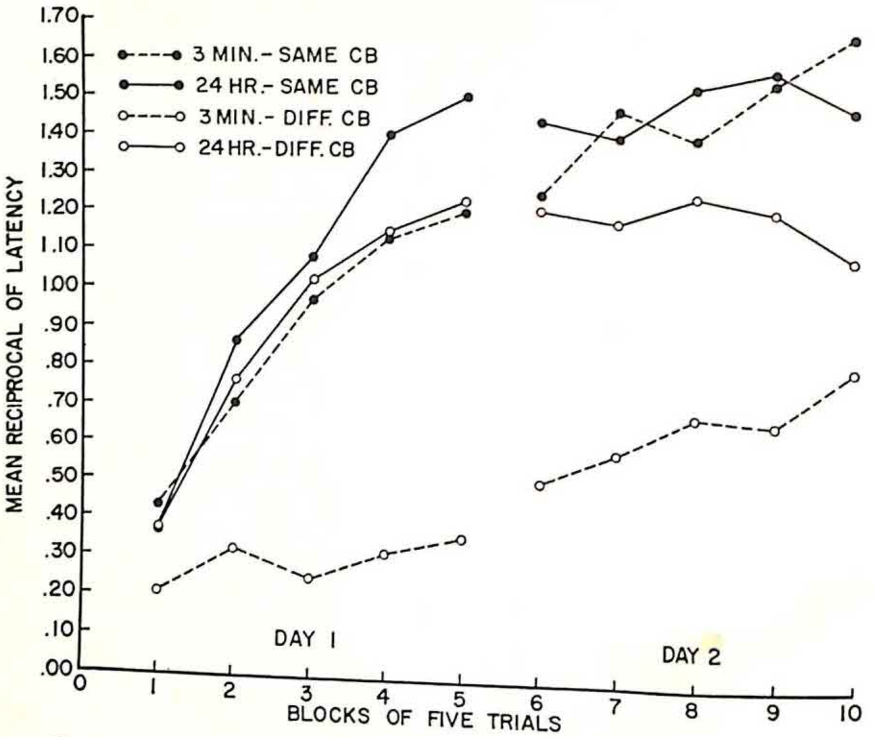


FIG. 1. Mean reciprocal of latency of hurdle jumping in blocks of five trials on Days 1 and 2 for the several groups.

ch of the groups. The improvement performance on Day 1 is clearly ident for all groups except Group 3 n.-Diff CB, whose performance ows no marked increase until Day 2. n Day 1, Group 24 hr.-Same CB is superior to all other groups, but on Day 2 its performance is equaled by Group 3 min.-Same CB.

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE OF RECIPROCAL OF LATENCY OF HURDLE JUMPING

Source	*df*	Day 1		Day 2	
		MS	*F*	*MS*	*F*
veen *S*s					
onditioning box (CB)	1	10.66	16.92***	24.25	12.12***
elay (D)	1	11.71	18.59***	6.10	3.05
B × D	1	4.22	6.70*	5.29	2.64
rror (b)	60	.63		2.00	
in *S*s					
ials (T)	4	5.51	91.83***	.26	2.60*
× CB	4	.59	9.83***	.06	.60
× D	4	.78	13.00***	.34	3.40*
× CB × D	4	.14	2.33	.03	.30
ror (w)	240	.06		.10	

$p < .05$.
$p < .001$.

The data for the 2 hurdle-jumping days were analyzed separately in order to avoid contamination of the postconditioning delay variable by the further delay occurring after Day 1. A trend analysis of variance of the Day 1 data was computed, with Conditioning Box (Same CB or Diff CB), Postconditioning Delay (3 min. or 24 hr.), and Trial Blocks as the factors. A summary table of the analysis appears in the left portion of Table 1. All factors except the triple interaction were significant. Because of the significant double interactions, the simple effects of the main factors were analyzed.

The interaction of the two between-*S*s factors, Conditioning Box and Postconditioning Delay, is illustrated in Fig. 2, where the grand means of

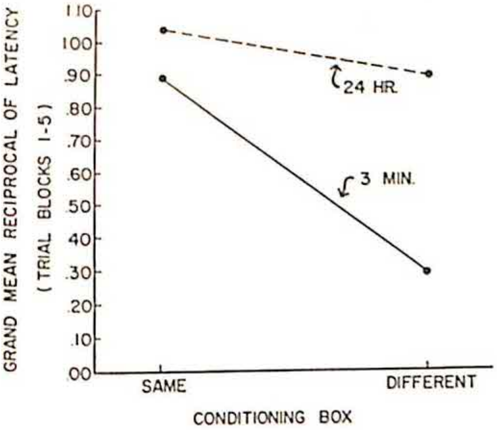


FIG. 2. Stimulus generalization gradients obtained with postconditioning delays of 3 min. and 24 hr.

Trial Blocks 1–5 are plotted. Since the abscissa represents stimulus differences in the boxes used for conditioning, the curves plotted are stimulus generalization gradients. Although the exact shapes of the gradients are unknown, they are drawn as linear for simplicity. On the assumption that the relationship is monotonic, it is clear that the gradient is steeper with a 3-min. than with a 24-hr. postconditioning delay. Using the error term from the main analysis, *t* tests indicated a significant stimulus generalization decrement for the 3-min. condition, $t\ (60) = 5.00$, $p < .001$, but not for the 24-hr. condition, $t\ (60) = 1.17$, $p > .20$. Also, there was a significant improvement in performance with the increase in postconditioning delay from 3 min. to 24 hr. when fear was conditioned in a box different from the start box of the hurdle-jumping apparatus, $t\ (60) = 5.08$, $p < .001$, but not when the same box was used, $t\ (60) = 1.25$, $p > .20$. A further analysis using only the data of Trial Block 5 yielded the same results except that the interaction term was significant at only the 6% level.

The stimulus dimension which underlies these generalization gradients consists in all likelihood of the several visual variations between the conditioning box and the start box of the hurdle-jumping apparatus. It might be thought that olfactory cues would be an important aspect of the stimulus complex to which fear was conditioned and, since different boxes were used, an important part of the generalization dimension. However, these cues were largely eliminated for all groups by the removal of the droppings from beneath the start box prior to hurdle jumping. That the remaining olfactory cues, those of the start box itself, did not importantly affect the results is indicated by a comparison of Group 3 min.-Same CB and the two 24-hr. groups. Such odor cues were not the same for these three groups because of differences in postconditioning delay and locus of conditioning, and yet they performed similarly. It will also be noted that recent odor cues were equal for the two 3-min. groups, as a result of running the *S*s in pairs, but they performed differently. The argument that the lowered performance of Group 3 min.-Diff CB still may be due to the removal from the stimulus complex during hurdle jumping of *S*'s own odor cues, which could be distinguished from those of another *S*, is also mitigated by the similarity in performance of the other three groups, for which this olfactory component of the stimulus complex obviously differed. Thus, it appears that olfactory cues do not contribute in any significant way to the results.

The significant within-*S*s interactions indicate that the trends of the performance curves on Day 1 depended upon both the Conditioning Box and the Delay variables. This conclusion is probably valid, because of the highly significant *F* values, despite the fact that the assumption of homogeneity of variance for the separate estimates of the within-*S*s error term was not tenable according to Bartlett's test, χ^2 (3) = 18.99, $p < .001$. Because of the interactions, the effect of Trials was evaluated separately for each group using a Trials × *S*s analysis of variance. The increases in performance shown in the curves of Fig. 1 were found, by these analyses, to be significant ($p < .001$) for all groups except Group 3 min.-Diff CB ($p > .10$).

The Day 2 data were analyzed with a trend analysis of variance in the same manner as for Day 1. A summary of the analysis appears in the right portion of Table 1. Of the between-*S*s factors, only the Conditioning Box main effect was significant. As may be seen in Fig. 1, the Same CB groups were superior to the Diff CB groups. Although a plot of the grand means on Day 2 for the several groups yielded stimulus generalization curves similar to those of Day 1, as seen in Fig. 2, the interaction term did not reach a satisfactory level of significance. Of the within-*S*s effects, Trials, and also its interaction with Delay, were significant. Further analyses, using the error term from the main analysis, indicated a significant Trials effect for the 3-min. condition, F (4, 240) = 5.10, $p < .001$, but not for the 24-hr. condition, F (4, 240) = .80, $p > .20$. As may be seen in Fig. 1, the performance of each 3-min. group improved over trials while that of each 24-hr. group remained at about the same level.

Discussion

The finding that postconditioning delay is an important variable affecting hurdle-jumping performance when different apparatus are used for conditioning and hurdle jumping is consistent with the hypothesis that the strength of the fear response to generalized stimuli increases over time. Thus, the failure to find learning of hurdle jumping immediately following conditioning in a previous study (McAllister & McAllister, 1962a) and in Group 3 min.-Diff CB of this experiment can be attributed to a stimulus generalization decrement of fear. The rapid learning which was found after a postconditioning delay of 24 hr. in the previous experiment and in Group 24 hr.-Diff CB of this study can be accounted for by the flattening of the stimulus generalization gradient of fear with time.

An alternative hypothesis might be proposed to account for the results without positing an increase in the fear response over time. During conditioning some other response (e.g., crouching) may have been learned to the stimuli of the situation and, when elicited during hurdle training, may have interfered with the learning of the hurdle-jumping response. With the further assumption that such a response decreases in strength with time, the difference in performance of the two Diff CB groups could be explained. However, such an explanation would predict a depression in the performance of Group 3 min.-Same CB also since the competing response should be elicited maximally in the situation in which it was originally learned. Since such a result was not obtained, the hypothesis is not convincing. Also, overt responses, which potentially could compete with hurdle jumping, were not observed to occur with any consistency.

Accepting, then, the hypothesis that the generalization gradient of fear flattens with time because of an increase in the fear response to generalized stimuli, there remains the question of why such a change should occur. Interpretations based simply on an increase in fear with time such as an incubation or consolidation of fear hypothesis (e.g., Bindra & Cameron, 1953) are not cogent since an increased response strength to the original, as well as to the generalized, stimuli would be predicted. This criticism applies also to explanations postulating a change over time of processes competing with the fear response or inhibiting it as well as to the notion that an increase in the strength of an emotional response with time is due to the loss of the adaptation of that response (Golin, 1961).

Any satisfactory explanation of the present data requires a provision for the differential effect of postconditioning delay along the stimulus dimension. One such interpretation assumes that the psychophysical scaling of the stimulus dimension would result in a different number of jnd's for the two delay conditions. If the procedures used in the scaling involved a short delay between presentations of the comparison stimuli, presumably more jnd's would result than with a longer delay. Thus, in the present study the original and generalized stimuli would be discriminable (separated by many jnd's) immediately following conditioning but not discriminable (separated by few, if any, jnd's) after a delay. As a consequence, little fear would be elicited by the generalized stimuli in the 3-min. condition but would be strongly elicited in the 24-hr. condition. Although this explanation is consistent with the data, it remains circular in the absence of independent evidence of the postulated difference in the psychophysical scale.

In spite of the lack of a satisfactory theoretical explanation, the empirical phenomenon would seem to have applicability to a variety of situations involving the use of generalized stimuli coupled with a temporal variable. To date, evidence for the flattening of the stimulus generalization gradient with time has, in addition to the present study, been reported in two quite dissimilar situations: one employing the running response in rats (Perkins & Weyant, 1958) and the other, the pecking response in pigeons (Thomas & Lopez, 1962). Both studies utilized responses learned on the basis of an appetitive need in contrast to the acquired need used in this experiment. In addition, the data of Mednick (1957) provide suggestive evidence for the same phenomenon with secondary generalization.

In one respect the present results differ from those of Perkins and Weyant and of Thomas and Lopez. Both of those studies reported a decrease over time in response strength to the training stimulus, although in the former study it was not significant. In contrast, an increase was found in this study. Although not significant, it should, perhaps, be given some consideration because of the findings on the second day of hurdle jumping. The performance of Group 3 min.-Same CB improved significantly on Day 2, after a 24-hr. delay, to equal that of Group 24 hr.-Same CB whose performance remained at about the same level as at the end of Day 1. Thus, in the present situation, it might be premature to dismiss the possibility that an increase in the strength of the fear response can occur over time even in the absence of stimulus generalization. If future research should indicate that such is the case, some of the explanations dismissed above might become plausible unless it is found that the differential effect of delay on responses to original and generalized stimuli still remains.

A question might be raised concerning the relatively poor performance on both days of hurdle jumping of Group 3 min.-Diff CB. The failure of that group to learn on Day 1 might be attributed to the absence of fear. However, if that is the reason, it would be expected on the basis of the present results that on Day 2, after the 24-hr. rest permitted an increase in the strength of the generalized fear response, performance would be comparable to that of Group 24 hr.-Diff CB on Day 1. Although the 3-min. group did learn on the second day, it was at a slower rate. The explanation of this result is likely that some fear, but not a sufficient amount to serve as the basis for the learning of hurdle jumping, was elicited on Day 1 and was extinguished as a result of the repeated exposures to the stimulus complex in the absence of shock. As a result, the amount of fear available on Day 2 was lessened and, hence, the learning was poorer.

REFERENCES

BINDRA, D., & CAMERON, L. Changes in experimentally produced anxiety with the passage of time: Incubation effect. *J. exp. Psychol.*, 1953, **45**, 197–203.

GOLIN, S. Incubation effect: Role of awareness in an immediate versus delayed test of conditioned emotionality. *J. abnorm. soc. Psychol.*, 1961, **63**, 534–539.

MCALLISTER, W. R., & MCALLISTER, D. E. Postconditioning delay and intensity of shock as factors in the measurement of acquired fear. *J. exp. Psychol.*, 1962, **64**, 110–116. (a)

MCALLISTER, W. R., & MCALLISTER, D. E. Role of the CS and of apparatus cues in the measurement of acquired fear. *Psychol. Rep.*, 1962, **11**, 749–756. (b)

MEDNICK, M. T. Mediated generalization and the incubation effect as a function of manifest anxiety. *J. abnorm. soc. Psychol.*, 1957, **55**, 315–321.

PERKINS, C. C., JR., & WEYANT, R. G. The interval between training and test trials as a determiner of the slope of generalization gradients. *J. comp. physiol. Psychol.*, 1958, **51**, 596–600.

THOMAS, D. R., & LOPEZ, L. J. The effects of delayed testing on generalization slope. *J. comp. physiol. Psychol.*, 1962, **55**, 541–544.

(Received July 30, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 583–586

EFFECT OF INSPECTION TIME AND DIRECTION OF ROTATION ON A GENERALIZED FORM OF THE SPIRAL AFTEREFFECT[1]

PAUL BAKAN AND KIYOE MIZUSAWA

Michigan State University

The spiral aftereffect, an apparent expansion or contraction of a spiral afterfixation of the rotating spiral, is produced in a more general form as an apparent expansion or contraction of a square after fixation of a rotating spiral. Duration of the spiral-square aftereffect is a positive function of time of inspection of the spiral. Duration of the aftereffect is longer for the apparently expanding aftereffect than for the apparently contracting effect. Duration of the aftereffect appears to be influenced by the nature of the postrotation stimulus upon which it is projected.

There has been considerable recent interest in the spiral aftereffect, with much of the work directed to studies of the relationship between the aftereffect and brain damage (Gallese, 1956; Holland & Beach, 1958; Philbrick, 1959; Price & Deabler, 1955; Spivack & Levine, 1957, 1959). Typically *S* fixates the center of a rotating spiral and then reports on the appearance of the spiral when it stops rotating. If the spiral appears to expand while rotating, the aftereffect is apparent contraction; if the spiral appears to contract while rotating, the aftereffect is apparent expansion.

The effect need not be confined to a spiral test stimulus, but can be demonstrated as a distortion of any structured visual stimulus. Though the generality of the aftereffect has been noted (Holland, 1957; Osgood, 1953, p. 243; Spitz, 1958a, 1958b) there have been no published reports of investigations of the spiral aftereffect in the more general form with nonspiral test stimuli. Failure to recognize the general nature of the spiral aftereffect has led to theoretical treatment of the effect limited to the special case where the same spiral is used as both inspection stimulus and test stimulus (George, 1953; Spitz, 1958a).

In its more general form the spiral aftereffect can be thought of as a type of figural aftereffect, since exposure to an inspection stimulus results in an apparent change in the appearance of a subsequently viewed test stimulus (Köhler & Wallach, 1944; Spitz, 1958a).

Two variables known to influence the duration of the spiral aftereffect are duration of inspection of the spiral, and direction of rotation of the spiral (Costello, 1960, 1961; Eysenck & Holland, 1960; Eysenck, Willett, & Slater, 1962; Holland, 1958; Holland & Eysenck, 1960; Spitz, 1958a; Spitz & Lipman, 1959; Truss & Allen, 1959). Duration of the aftereffect is positively related to duration of inspection of the spiral. Rotation of the spiral which produces apparent contraction and an aftereffect of apparent expansion produces a longer aftereffect than rotation which produces apparent expansion and a contraction aftereffect. If the effect of a rotating spiral on sub-

---

[1] This paper was written while the senior author was a visiting worker at the Applied Psychology Research Unit, Cambridge, England. It was supported in part by a grant from the National Science Foundation.

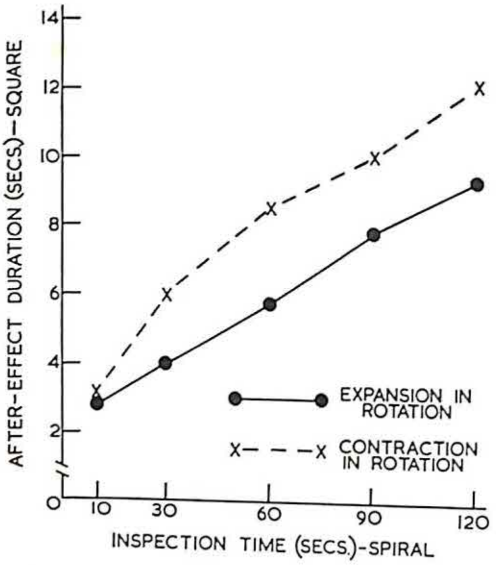


FIG. 1. Duration of aftereffect on square as a function of inspection time on spiral for two directions of spiral rotation.

sequent perception of the stationary spiral is a special case of the more general effect of a rotating spiral, then it would follow that the inspection time and direction of rotation should influence the duration of the general spiral aftereffect just as they influence the duration of the special case of the effect. This hypothesis is tested in the present study, where a white square, rather than a spiral, is used as the test figure.

METHOD

Design.—A sample of 75 undergraduates was randomly divided into five groups of 15. Groups differed with respect to inspection time, the inspection times being either 10, 30, 60, 90, or 120 sec. Independent groups were used to avoid the possibility of confounding residual aftereffects. Four measures of duration of aftereffect were taken for each *S*; two measures were taken after clockwise rotation and two after counterclockwise rotation. Order of rotation was counterbalanced among *S*s.

Stimuli.—The spiral inspection figure measured 7.25 in. in diameter with a single black throw of 2.5 revolutions on a white background. The fixation point was a standard acorn nut which fastened the center of the disc to a driveshaft. The spiral, powered by batteries has a rotation speed of 82 rpm.[2]

The test figure on which the aftereffect was projected was a white paper square, 6 × 6 cm., mounted on a gray background. The *S* fixated a black dot in the center of the test figure.

Procedure.—The *S*s were tested individually. The *S* was seated in front of a reduction screen with one eye about 1 in. in front of a hole 8 mm. in diameter, through which he viewed the stimulus from a distance of 8 ft. All *S*s were given two practice trials with 30 sec. inspection, one in each direction of rotation. There followed four test trials with inspection time the same on each trial. Each set of four trials consisted of two successive trials in each rotational direction with trials separated by 2 min. On each trial *S* viewed the test square immediately after viewing the rotating spiral, and reported to *E* upon the cessation of apparent change in the test square.

RESULTS

The mean duration of the aftereffect for the five groups, differing in length of inspection period, is shown in Fig. 1. The two functions shown represent functions for the two directions of rotation. Clearly the function relating duration of inspection to duration of aftereffect is positively monotonic and practically linear

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF AFTEREFFECT DURATION

Source	*df*	*F*
Inspection time (I)	4	14.6***
Between-*S*s in same I (*S*)	70	(76)
Rotation direction (D)	1	29.8***
I × D	4	1.3
I × D × *S*	70	(22)
Total	149	

Note.—Numbers in parentheses are *MS*s.
*** $p = .001$.

---

[2] The apparatus was supplied by Psychological Research and Development Corporation, Boston, Massachusetts.

within the range explored. An analysis of variance summarized in Table 1 shows the effect of inspection time significant at the .001 level.

The effect of direction of rotation is also clear. When the spiral appears to contract in rotation, the aftereffect of apparent expansion lasts longer than the aftereffect of apparent contraction produced by a spiral which appears to expand in rotation. The difference due to direction of rotation is significant at the .001 level.

DISCUSSION

The similar effect of inspection time and direction of rotation on duration of the specific and the general spiral aftereffect supports the hypothesis that the effects induced by a rotating spiral are of the same general nature whether projected upon the same spiral or on any other textured surface. There is nevertheless a puzzling aspect in the data of this experiment, namely that the duration of the aftereffect when projected on the square is less than durations reported in experiments where the inspection figure and test figure are the same spiral.

The difference in duration of aftereffect suggests that the test figure upon which the aftereffect is projected may influence the duration of the aftereffect, even if it does not determine the existence of the effect. The duration of the aftereffect could be a joint function of a state of the brain induced by the rotating spiral and a state induced by the specific test figure.

Two studies support this view. Costello (1960) used as an inspection figure an 8.24-in. spiral viewed at 6 ft.; the test figure was a photograph of this spiral reduced in diameter and viewed at a shorter distance to produce the same visual angle. After a 1-min. inspection period he found the duration of the aftereffect less with the photograph than with the full size spiral. Spivak (1961) found the duration of aftereffect smaller when projected on a white square than when projected on the stationary spiral itself. Further research leading to knowledge of the properties of the test figure which influence the duration of the aftereffect is indicated. The existence of such an influence means that any adequate theoretical account of these effects will have to allow for the interaction of the effects of both the inspection and test figure.

The difference in duration of the spiral aftereffect due to direction of rotation is a finding noted previously. The present study extends this finding to the general case of the spiral aftereffect. The reason for this difference is not clear. A factor which may help to account for the difference is a differential ease of fixation between the inward and outward rotation. It is suggested that inward rotation facilitates fixation at the center of the spiral since movement is directed toward the fixation point. With outward rotation there is apparent movement away from the center which may induce the eye to follow the movement and thus impair fixation, thus resulting in less effective satiation.

REFERENCES

COSTELLO, C. G. Further observations on the spiral after-effect. *Percept. mot. Skills*, 1960, **11**, 324.

COSTELLO, C. G. Massed practice on the spiral after-effect and the homeostatic nature of excitation-inhibition. *Percept. mot. Skills*, 1961, **12**, 11–14.

EYSENCK, H. J., & HOLLAND, H. C. Length of spiral after-effect as a function of drive. *Percept. mot. Skills*, 1960, **11**, 129–130.

EYSENCK, H. J., WILLETT, R. A., & SLATER, P. Drive, direction of rotation, and massing of practice as determinants of the duration of the aftereffects from the rotating spiral. *Amer. J. Psychol.*, 1962, **75**, 127–133.

GALLESE, A. J. Spiral after-effect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.

GEORGE, F. A. On the theory of the figural after-effect. *Canad. J. Psychol.*, 1953, **7**, 167–171.

HOLLAND, H. C. The Archimedes spiral. *Nature*, 1957, **179**, 432–433.

HOLLAND, H. C. Some determinants of seen after-movements in the Archimedes spiral. *Acta Psychol., Amsterdam*, 1958, **14**, 215–222.

HOLLAND, H. C., & BEACH, H. R. The spiral after-effect as a test of brain damage. *J. ment. Sci.*, 1958, **104**, 466–471.

HOLLAND, H. C., & EYSENCK, H. J. Spiral after-effect as a function of length of stimulation. *Percept. mot. Skills*, 1960, **11**, 228.

KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

OSGOOD, C. E. *Method and theory in experimental psychology.* N. Y.: Oxford, 1953.

PHILBRICK, E. B. The validity of the spiral after-effect as a clinical tool for diagnosis of brain pathology. *J. consult. Psychol.*, 1959, **23**, 39–43.

PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of the spiral after-effect. *J. consult. Psychol.*, 1955, **19**, 299–302.

SPITZ, H. H. Neural satiation in the spiral after-effect and similar movement after-effects. *Percept. mot. Skills*, 1958, **8**, 207–213. (a)

SPITZ, H. H. The present status of the Köhler-Wallach theory of satiation. *Psychol. Bull.*, 1958, **55**, 1–28. (b)

SPITZ, H. H., & LIPMAN, R. S. Some parameters in the perception of the spiral after-effect. *Percept. mot. Skills*, 1959, **9**, 81.

SPIVAK, A. P. Extraversion-introversion and spiral after-effects. Unpublished master's thesis, Michigan State University, 1961.

SPIVACK, G., & LEVINE, M. The spiral after-effect and reversible figures as measures of brain damage, and memory. *J. Pers.*, 1957, **25**, 767–778.

SPIVACK, G., & LEVINE, M. Spiral after-effect and measures of satiation in brain injured and normal subjects. *J. Pers.*, 1959, **27**, 211–227.

TRUSS, C. V., & ALLEN, R. M. Duration of the spiral after-effect in cerebral palsy: An exploratory study. *Percept. mot. Skills*, 1959, **9**, 216–218.

(Received August 1, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 587–592

STIMULUS GENERALIZATION AS A FUNCTION OF TESTING PROCEDURE AND RESPONSE MEASURE[1]

RICHARD H. HISS[2] AND DAVID R. THOMAS

Kent State University

36 homing pigeons received 10 days of VI training with a CS of 550 mμ and then were tested for generalization in extinction. Group 1, Group 2, and Group 3 ($n = 9$ each) were run to a criterion of extinction at 550, 530, and 510 mμ, respectively. Group 4 ($n = 9$) received a random presentation of the 3 stimuli for 1 hr. For all Ss during both training and testing 30-sec. stimulus-on periods were alternated with blackout periods of 10 sec., thus creating discrete trials. The measures employed were: number of responses on the 1st test trial, total responses to extinction (or during the test), latency of the 1st response during the 1st 5 trials to which responding occurred, and rate of responding during the 1st 5 trials to which responding occurred. For each measure a gradient was obtained from Group 4 and from the combination of Groups 1, 2, 3, each of which contributed 1 point to the curve. It was found that the comparison of stimuli during testing is neither essential for the creation of a reliable gradient nor is it effective in altering the resulting gradient in any consistent way.

A major purpose of this study was to compare two different testing procedures for obtaining wave length generalization gradients after training pigeon *S*s to respond to a single stimulus. The first procedure may be called noncomparison testing. It consists of presenting a single stimulus to each *S* and measuring the resulting response strength. A separate group is tested at each of a number of different stimulus values. Studies using this method include: Grice and Saltz (1950), Kling (1952), and Margolius (1955). The other procedure involves successively presenting a number of stimuli to each *S* and recording the response strength to each. Each *S* acts as its own control and complete gradients may be obtained from individual *S*s. This procedure, which may be called comparison testing, has been used extensively by Guttman and Kalish (1956) and their associates (Hanson, 1957; Honig, Thomas, & Guttman, 1959; Thomas & King, 1959).

The importance of comparison of stimuli in generalization was suggested by Lashley and Wade (1946). They argued that comparison of stimuli *along the same dimension* is essential in order to create a gradient of generalization. Most studies which have attempted to test their hypothesis have dealt with the nature of comparisons occurring during training, (e.g. Jenkins & Harrison, 1960). The major concern of the present study was to investigate the role that comparison of stimuli during testing may play.

It has been argued by some (e.g. Grant & Schiller, 1953; Prokasy &

---

[1] This research was supported in part by United States Public Health Service Research Grants M-4203 and MH-04203-02 to the second author. The study was carried out by the first author under the direction of the second in partial fulfillment of the requirements for the MA degree at Kent State University. A preliminary report of these findings was made at the Midwestern Psychological Association meetings in Chicago, May 1962.

[2] Now at Syracuse University.

Hall, 1961) that the best measure of generalization is that taken on the first test trial. In support of their argument they refer to studies in which generalization decrements develop during the course of testing (e.g. Wickens, Schroder, & Snide, 1954). The present study was designed to indicate whether the opportunity to compare stimuli during testing alters the generalization gradient in the operant free-responding situation. Two comparisons will be of interest in this regard, one between comparison and noncomparison gradients and another between comparison gradients obtained on the first test trial and those obtained thereafter.

An additional purpose of this study was to compare generalization gradients obtained with a number of different response measures. It was anticipated that the effects of comparison of stimuli, for example, might depend on the response measure used. Since most studies of stimulus generalization have reported only one response strength measure it is impossible to say what differences in findings might have resulted from the use of a number of them. In order to maximize the number and diversity of response measures in the present study it was decided to incorporate the features of the operant free-responding situation with those of a discrete-trial procedure.

METHOD

Subjects.—The *S*s were 36 experimentally naive homing pigeons maintained at 75% of their ad lib. weight.

Apparatus.—Two identical Skinner type key pecking apparatus were used. Each box had the following internal dimensions: width, 10 in.; depth, 12 in.; and height, 12 in. Walls and ceilings were painted flat black; floors were heavy gauge wire mesh. The *S*'s key of translucent plastic was exposed through a $\frac{7}{8}$-in. circular opening placed 8 in. above the floor on one wall of the box. Directly below the key in the wall of each box was a $1\frac{1}{4}$-in. square aperture through which *S*s had access to food for 3 sec. during each cycle. Between cycles the food was lowered beyond *S*'s reach. The magazine containing the food was illuminated with a 6-v. bulb which was lighted whenever the food magazine was raised. Aside from the stimulus light and the magazine light during the magazine cycles, the boxes were in darkness throughout the experiment.

The stimuli were provided by Bausch and Lomb interference filters. The training stimulus (CS) was provided by a filter with a transmission peak at 550 mμ. Testing stimuli were provided by filters with transmission peaks at 550, 530, and 510 mμ. The stimuli were illuminated by Sylvania projection bulbs. Throughout the experiment a masking noise was supplied to the boxes by a Grason-Stadler noise generator, Model 901.

Procedure.—Upon arrival at the laboratory, all *S*s were weighed, individually housed, and allowed free access to food and water. The weight of all *S*s was measured on the day of arrival and each day thereafter. Food deprivation was not begun until a stable weight level was achieved by each *S*. After 10 days a stable weight level was obtained and food deprivation was begun. When *S*s reached 75% of their ad lib. weight training was begun.

All *S*s were adapted, magazine trained, and conditioned to peck according to a set schedule covering 5 days. Then *S*s were placed on successively increasing fixed ratios of 15, 30, 60, and finally on a variable-interval schedule. The mean interreinforcement interval was about 1 min.; the range was from 4 sec. to 4 min. During VI training a shutter was used which alternated 10-sec. periods of blackout with 30-sec. periods of stimulus presentation. Reinforcements were never given during blackouts and during training sessions the key was illuminated with 550 mμ.

VI training was terminated at the end of 10 daily hour-long sessions, providing that a criterion of stable responding had been reached. After the eighth day of VI training an interval of ±20% of total responses was established. If the number of responses made on VI Days 9 and 10 fell within this interval, *S* was tested for generalization following training on Day 10. If the number of responses fell outside of this interval a new interval was determined and applied to the two subsequent training sessions. The *S*s required a median of 11 days of VI training to achieve criterion.

Following the last session of VI training *S*s were assigned to four groups and were immediately tested for generalization. Group 1 ($n = 9$) was exposed to 550 mμ, Group 2 ($n = 9$) was exposed to 530 mμ, and Group 3 ($n = 9$) was exposed to 510 mμ. These *S*s were placed on extinction and were run to a criterion of 10 successive presentations of the stimulus with a total of no more than 10 responses. Group 4 ($n = 9$) received 90 random presentations of the 3 stimuli 550, 530, and 510 mμ, 30 of each. Groups 1, 2, and 3 will be called the noncomparison groups and Group 4 will be called the comparison group. For all groups during generalization testing both the number of responses in each 30-sec. stimulus presentation and the latency of the first response during that period were recorded.

Results and Discussion

With noncomparison testing a gradient of number of responses to extinction does exist. The median number of responses made by the *S*s in each group are: 550 mμ, 2237; 530 mμ, 1077; and 510 mμ, 136. These differences are significant according to the Kruskal-Wallis one-way analysis of variance, $H(2) = 8.4$ ($p < .02$). The use of a nonparametric analysis of these data was dictated by the atypical scores of a few *S*s. For the sake of uniformity, nonparametric analyses were used throughout the study. The median number of responses made by each of the 9 *S*s in the comparison group to each of the three stimuli are: 550 mμ, 533; 530 mμ, 42; and 510 mμ, 3. This gradient was analyzed with the Friedman two-way analysis of variance by ranks, $\chi_r^2\ (3, 9) = 18.0$ ($p < .01$).

Although the noncomparison groups differed in number of responses to extinction, they did not differ significantly in the time required to achieve the extinction criterion. The 550-mμ group took a median of 49 min. compared to 53.5 min. in the 530-mμ and 45 min. in the 510-mμ group. The fact of a significant gradient in number of responses to extinction but not in time required suggests that relative differences in response rate to the different test stimuli are maintained throughout the process of extinction. In the original Guttman and Kalish (1956) article gradients obtained during the four quarters of the generalization test in extinction were compared and no evidence of steepening of the gradient was found. A similar analysis of the comparison group data in the present experiment corroborated this finding. In the present experimental situation it may therefore be concluded that the testing for generalization, even over a prolonged session of extinction, does not distort the resulting gradient.

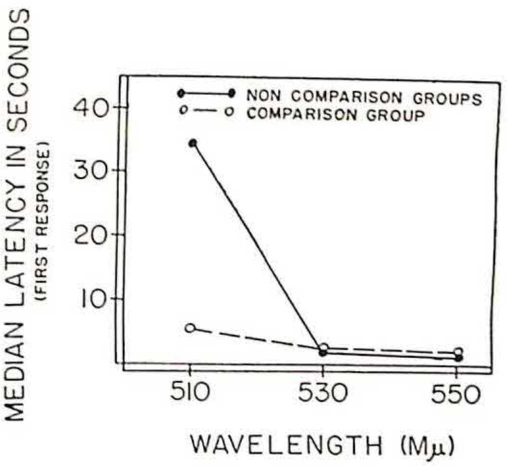


FIG. 1. Median latency in seconds (first response)—noncomparison groups and comparison group.

It is of considerable theoretical interest to observe generalization gradients as obtained on the first test trial. In the present experiment this means the first 30 sec. presentation of the stimulus. Note that for the comparison group the first exposure to, say, 550 mμ may have been preceded by an exposure of 510 mμ and 530 mμ.

Figure 1 presents data on the median latency on the first test trial. Latency scores greater than 30 sec. were possible when *S*s did not respond

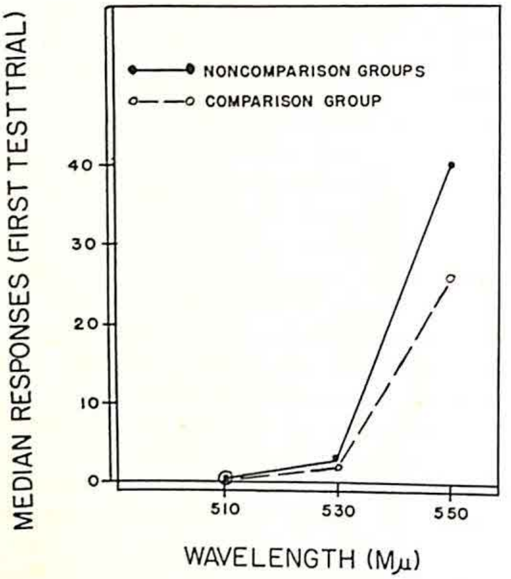


FIG. 2. Median responses (first test trial) —noncomparison groups and comparison group.

during the first test trial. The noncomparison gradient is significant, H (2) = 9.6 ($p < .01$), however, the comparison gradient is not significant, χ_r^2 (3, 9) = 1.2 ($p > .05$).

A statistical comparison between gradients obtained under comparison and noncomparison conditions presents certain methodological difficulties. In order to compare the slope of such gradients the noncomparison data were treated as follows: Random triplets were selected with one value from each of the three stimulus conditions. These triplets were then treated as if they had come from the same animal. This resulted in an error on the conservative side since the variability within the triplets so determined was probably greater than if the same *S*s had been used. The percent of the total time (or number of responses) given to the CS (550 mμ) was taken as an index of the steepness of the gradient. Percent values were also obtained from the comparison group and a Mann-Whitney test was used to compare the two sets of percent scores. A similar analysis was used in all comparison vs. noncomparison analyses, except that the random triplets were separately determined for each analysis.

For the data reported in Fig. 1 the Mann-Whitney test revealed a difference in slope which is significant at the .01 level. Thus, for this measure, comparison flattens the gradient.

In Fig. 2 the data presented are the median number of responses made on the first test trial. The comparison and noncomparison gradients are both significant: H (2) = 17.32 ($p < .01$) and χ_r^2 (3, 9) = 7.39 ($p < .03$). For the two gradients the Mann-Whitney test reveals a difference in slope which is significant beyond the .01 level. Again the comparison group gradient was the flatter of the two.

In the present study most *S*s in the 510-mμ condition and some *S*s in the 530-mμ condition failed to respond at all during the first stimulus presentation, therefore medians of measures taken on the first test trial may give a distorted picture of typical performance. It would be interesting to inquire what the gradients would be like if only trials on which *S* responded

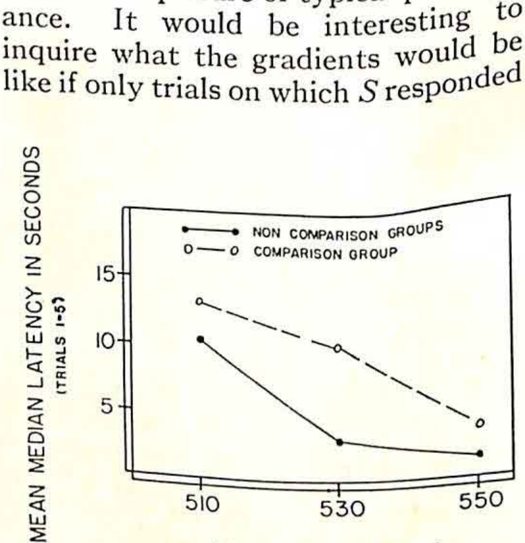


FIG. 3. Mean of the median latencies in seconds for Trials 1–5, noncomparison groups and comparison group.

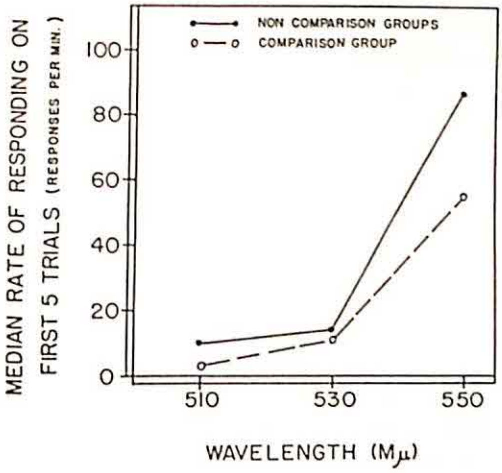


FIG. 4. Median rate of responding for Trials 1–5, noncomparison groups and comparison group.

were considered. Data on latency and rate for the first five trials on which responding occurred are presented in Fig. 3 and 4. The latencies in Fig. 3 are the mean of the median latencies for the 9 *S*s in each of the noncomparison groups and the mean of the median latencies to each stimulus in the comparison group. An analysis of these data shows: for the comparison group, χ_r^2 (3, 9) = 4.22 (*ns*), and for the noncomparison groups, *H* (2) = 10.8 ($p < .01$). There is no significant difference between the steepness measures for these two gradients.

The data in Fig. 4 represent the median rate of responding on the first five trials (2.5 min.) to which *S*s responded. Both gradients are statistically significant: comparison group, χ_r^2 (3, 9) = 18.0 ($p < .01$), and the noncomparison groups, *H* (2) = 13.2 ($p < .01$). This indicates that in addition to the fact that *S*s are less likely to respond (at all) to 530 mμ and 510 mμ, when they do respond to these stimuli they do so at a lower rate and with a longer latency. Again the two gradients fail to differ significantly in steepness.

The following conclusions may now be offered:

A reliable gradient may be obtained with a variety of response measures following either comparison or noncomparison testing procedures. The comparison of test stimuli during generalization testing is therefore not essential for the establishment of the gradient. Further, there is no evidence that opportunity for such comparison alters the resulting gradient in any consistent way. The measures of latency of the first response and number of responses on the first test trial yielded steeper gradients for the noncomparison than the comparison procedure. It should be remembered that for the comparison group the first test trial at a given stimulus might be preceded by test trials to one or two other values. This opportunity for comparison did not steepen the gradient (as might be predicted from a discrimination learning hypothesis) but rather flattened it. The measures of median latency and median responses per minute calculated over the first five trials to which *S* responded showed no slope differences attributable to testing procedure. The finding of gradients with these two measures is particularly interesting because it indicates that in addition to the fact that *S*s are less likely to respond to non-CS stimuli than to the CS, when they do respond to such stimuli, they do so at a lower rate and with a longer latency.

Although a reliable gradient of responses to extinction was obtained with the noncomparison groups, no gradient of time to extinction was found. This suggests that generalization testing may be carried out extensively in extinction without any distortion in the relative response strengths to the different test stimuli.

REFERENCES

GRANT, D. A., & SCHILLER, J. J. Generalization of the conditioned galvanic skin responses to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.

GRICE, G. R., & SALTZ, E. The generalization of instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.

Guttman, N., & Kalish, H. L. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hanson, H. M. Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*, 1957, **125**, 888–889.

Honig, W., Thomas, D. R., & Guttman, N. Differential effects of continuous extinction and discrimination training of the generalization gradient. *J. exp. Psychol.*, 1959, **58**, 145–152.

Jenkins, H. M., & Harrison, R. H. Effect of discrimination training on auditory generalization. *J. exp. Psychol.*, 1960, **59**, 246–253.

Kling, J. W. Generalization of extinction of an instrumental response to stimuli varying on the size dimension. *J. exp. Psychol.*, 1952, **44**, 339–346.

Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.

Margolius, G. Stimulus generalization of an instrumental response as a function of the number of reinforced trials. *J. exp. Psychol.*, 1955, **49**, 105–111.

Prokasy, W. F., & Hall, J. K. Stimulus generalization: Lashley and Wade revisited. Paper read at the Psychonomic Society meeting, New York, September 1961.

Thomas, D. R., & King, R. A. Stimulus generalization as a function of level of motivation. *J. exp. Psychol.*, 1959, **57**, 323–328.

Wickens, D. D., Schroder, H. M., & Snide, J. D. Primary stimulus generalization of the GSR under two conditions. *J. exp. Psychol.*, 1954, **47**, 52–56.

(Received August 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 593–596

GROUP SIZE AND NUMBER OF VICARIOUS REINFORCEMENTS IN VERBAL LEARNING[1]

ALBERT R. MARSTON AND FREDERICK H. KANFER
University of Wisconsin *University of Oregon Medical School*

The relationship of group size to vicarious reinforcement (VR) was investigated under conditions controlled either for number or percentage of reinforced responses heard by Ss in simulated groups. The effect of adding direct reinforcement (DR) to the observing S was examined at each level of group size. In a modified verbal conditioning task, with "Good" as reinforcer and Humans as critical class, Ss said words intermittently with tape recorded voices. It was found that decreasing proportion of reinforced vicarious responses by increasing group size resulted in significantly lower learning. When number of reinforcements increased (proportion held constant), learning curves did not differ. Adding DR to VR did not significantly increase learning at any level of group size.

In an earlier study investigating vicarious reinforcement (VR) in verbal learning with simulated groups (Kanfer & Marston, 1963), it was found that the presence of VR significantly increased learning but that the addition of direct reinforcement (DR) to the observing *S* had no further effect. It appeared possible that the size of the observed group (nine members) resulted in the relative ineffectiveness of DR by limiting the proportion of the total responses made by the *O*. Earlier studies have investigated VR either with one observed responder (e.g. Lewis & Duncan, 1958) or with a larger group (e.g. McBrearty, Marston, & Kanfer, 1961) but without varying size. The purpose of the present study was to investigate the relationship of size of observed group to amount of learning in the observer, with controls for both DR and VR. The relationship of research in VR to various types of organized group learning, whether in the formal classroom or in such less structured situations as group psychotherapy, makes the group size variable particularly relevant.

Group size can be varied in two ways with respect to controlling the number of reinforced correct responses given by a simulated group. Each additional group member can add a fixed number of correct responses (CR), thus increasing number of CRs with increasing group size and keeping the proportion of CRs to total number of observed responses constant. Or, one can keep the number of CRs fixed, distributing them among the available members. In the latter procedure, the proportion of reinforcement would decrease as group size increases. These alternate ways of varying group size, compared in the present study, reflect the implication of the group size variable for the general problem of the relationship of VR to direct reinforcement. Changing group size cannot be accomplished

---

[1] This study was supported in part by Research Grant MH 06922-01 to the second author, from the National Institute of Mental Health, United States Public Health Service. The authors wish to acknowledge the assistance of D. Leventhal and D. Slavin in the collection and analysis of the data.

without varying either number or proportion of VR, and conversely, one of the critical means for varying number of reinforcements in the VR procedure is the manipulation of group size.

METHOD

Subjects and apparatus.—One hundred twenty-three male undergraduates served as *S*s. The *S*s were seated at a table in a darkened room. At eye level on a 2 × 3 ft. black board in front of them was mounted a signal light. To the left of the light a microphone rested in a slot on the board. Throughout the experiment *S* wore a set of earphones. The earphones and microphone were part of an intercommunication system with the next room where *E* recorded responses. A Wollensak (Model T-1500) tape recorder served as an amplifier in the system. With his own microphone and earphones, *E* was able to hear *S* and speak to him, as well as to play the tape recorded "group" responses into both earphones. The *E* also used a telegraph key to activate the light on *S*'s board.

Procedure.—Simulated groups were tape recorded by male graduate students who read prepared lists of words taken from earlier group conditioning protocols. The original group conditioning procedure consisted of a modification of individual verbal conditioning in which *S*s in a group said words in turn. The *E* stood in front of the group and reinforced words from a selected class by saying "Good." In all, six tapes were recorded, varying in number of voices on the tape and number of reinforced correct responses (Human nouns) on the tape. Following each correct response on the tape the word "Good," spoken by one of the authors, was recorded. On all tapes 50 "rounds" of responses were recorded. A round consisted of a series of responses with each voice having one turn. Regardless of the *total* number of correct responses (CR) on a tape, the number per block of 10 rounds increased over the five blocks. This increase simulated a linear learning function progressing at a 45° angle from the origin.

The *S* reported to a hall where the experiment was to be conducted to find a row of office doors containing "Experiment in Progress" signs. He was ushered into the experimental room and asked to wait a few moments until *E* had settled the other (nonexisting) *S*s. When *E* returned, he instructed *S* to say a word whenever his signal light flashed. He also told him that other *S*s in other rooms would be performing the same task in turns and that he would be able to hear some of what was going on in the other rooms through the earphones which he was to wear throughout the experiment (for complete instructions see Kanfer & Marston, 1963).

The *E* began the conditioning procedure for all *S*s by activating the signal light 10 times at approximately 12-sec. intervals. These 10 trials, with no treatment administered, constituted a measure of operant or base level frequency of Human nouns.

After the last operant response, *E* pressed the "play" button of the recorder and allowed *S* to hear the first round of responses of his simulated group. After the last taped response in the round or after the subsequent "Good" if a Human noun were given, the "Stop" button was pushed, and *S* was signaled to respond. This sequence continued for 50 rounds of trials. At the end of the experiment, *S* filled out the following questionnaire:

1. What do you think was the purpose of the experiment?
2. Did you think it was easy or difficult? (Why?)
3. How did you go about choosing your words?
4. What did you hear through the earphones?
5. Did what you hear affect your behavior? (How?)

Responses on the questionnaire were rated for awareness of response-reinforcement contingency on a 6-point scale. Only 3 *S*s were aware of the contingency (rating of 5 or 6); these were eliminated from the analyses. Since, in an earlier study (Kanfer & Marston, 1963) *S*'s awareness that he was hearing a tape recording did not affect his learning, this awareness was not evaluated in the present study.

Experimental design.—A 3 × 2 × 2 × 6 factorial design was used. The *first* factor was group size; the tapes contained one, three, or five voices. The *second* factor concerned the relationship of number of reinforced responses on the tape to group size. At one level, as group size increased the proportion of reinforced responses on the tape remained constant (.30), thus increasing the number of reinforced responses (15 for Size 1, 45 for Size 3, and 75 for Size 5). In these groups each added voice to the tape contributed another unit of 15 reinforced responses. At the second level, the number of reinforced responses remained constant (30) and equally

TABLE 1

SUMMARY OF THE EXPERIMENTAL DESIGN, SHOWING PROPORTION OF REINFORCED RESPONSES ON THE TAPE IN EACH OF 12 GROUPS

VR Proportions	VR Group Size			VR + DR Group Size		
	1	3	5	1	3	5
Constant	.30	.30	.30	.30	.30	.30
Changing	.60	.20	.12	.60	.20	.12

distributed over all voices, thus decreasing the proportion of reinforced responses as group size increased (.60 for Size 1, .20 for Size 3, and .12 for Size 5). The *third* factor compared vicarious reinforcement alone with vicarious reinforcement plus direct reinforcement to the live *S* whenever he gave a CR. The six repeated measurements on each *S* (one operant and five blocks of 10 trials each in acquisition) constituted the fourth factor. The three between-*S*s factors provided 12 groups of 10 *S*s each. Table 1 summarizes the design.

RESULTS AND DISCUSSION

Table 2 presents a summary of the analysis of variance on frequency of CRs for each of six blocks of trials. The primary empirical question concerns the effectiveness of VR as a function of number and proportion of reinforcements. In Table 2, the variable labeled *Reinforcement* refers to the two different procedures of varying vicarious reinforcements when varying size of simulated group (constant and changing proportion). Because the study was primarily concerned with the interactive effect of group size and proportion of VR, the main effects of group size and reinforcement, which combined groups of varying proportions of VR, were of secondary interest. The significant Size × Reinforcement interaction ($F = 15.76, p < .001$) is illustrated in Fig. 1. As indicated by a test between groups in which the proportion of reinforcement was kept constant, these learning curves did not differ in overall height ($F < 1.0$). However, when proportion of vicarious (taped) CRs increased with decreasing group size the overall height of the curves differed significantly ($F = 28.37$,

TABLE 2

ANALYSIS OF VARIANCE ON FREQUENCY OF CRITICAL RESPONSES IN BLOCKS OF 10 TRIALS EACH

Source	*SS*	*df*	*MS*	*F*
Reinforcement (A)	22.54	1	22.54	6.21*
VR vs. VR + DR (B)	2.15	1	2.15	
Size (C)	92.20	2	46.10	12.70***
A × B	1.75	1	1.75	
A × C	114.39	2	57.20	15.76***
B × C	3.14	2	1.57	
A × B × C	12.80	2	6.40	1.76
*S*s in groups	391.71	108	3.63	
Blocks (D)	272.94	5	54.59	40.14***
A × D	12.14	5	2.43	1.79
B × D	3.01	5	0.60	
C × D	68.44	10	6.84	5.03***
A × B × D	1.98	5	0.40	
A × C × D	115.99	10	11.60	8.53***
B × C × D	9.99	10	1.00	
A × B × C × D	17.83	10	1.78	1.31
D × *S*s in groups	737.21	540	1.36	

* $p < .05$.
*** $p < .001$.

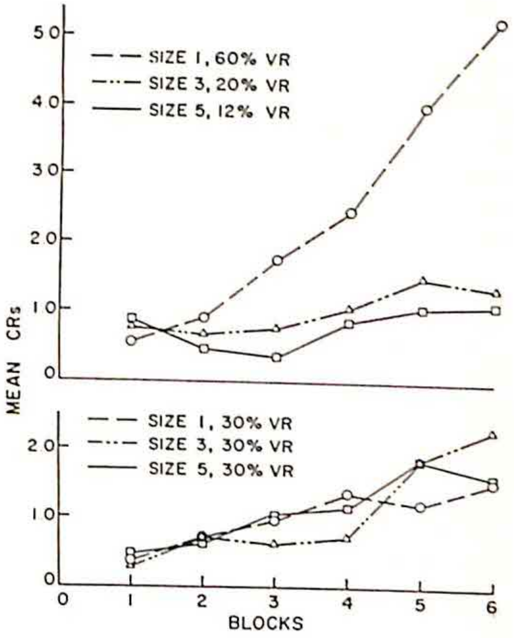


FIG. 1. Comparison of the learning functions of groups varying in size, shown separately for the constant proportion of VR and changing proportion of VR conditions.

$p < .001$). As can be seen in Fig. 1, the groups with constant proportion of CRs on the tape did not differ in slope; whereas the groups varying in proportion of CRs showed the steepest slope for the group with the highest proportion (Reinforcement × Size × Blocks $F = 8.53$, $p < .001$). Regardless of taped group size all of the groups with 30% VR increased significantly over blocks (Fs = 2.88, 9.43, and 4.36 for Sizes 1, 3, and 5, respectively). In the groups receiving varying proportions only the group with 60% CRs on the tape increased significantly over blocks ($F = 49.23$, $p < .001$). There was, however, significant learning when all groups were combined (F for Blocks $= 40.14$, $p < .001$).

The main effects become more meaningful in light of the interactions described above. The main effect of *Reinforcement* ($F = 6.21$, $p < .025$) reflects the overall superiority of the groups receiving varying proportions of VR. Although the average proportions of CRs on the tapes of these groups was only slightly higher (30.67% vs. 30.00%) the high rate of learning in the 60% group more than offset the failure to learn in the 20% and 12% groups. Also, the significant main effect for Size ($F = 12.70$, $p < .001$) supports the simple main effects reported previously, with Size 1, containing the 60% VR group, showing the greatest amount of learning. The interaction of Size × Blocks ($F = 5.03$, $p < .001$) indicates the superiority of Size 1 in rate as well as overall learning. These results show that though there is a close relationship between proportion of taped CRs and amount of learning, this relationship is not a constant ratio; i.e. the average amount of learning does not correspond directly to the average proportion of VR.

The results indicate that the effect of vicarious reinforcement follows principles well established in the literature of direct reinforcement. Estes (1956) pointed out that all reinforcement theories agree that strength of habit is increased by reinforced trials and decreased by nonreinforced trials. From this it follows that, "habit strength . . . should be directly related to percentage of reinforcements during a fixed number of trials" (p. 16). Numerous studies have empirically demonstrated this principle (e.g. Grant, Hake, & Hornseth, 1951). The relationship of VR to learning also appears dependent upon percentage of reinforcements. In fact, as can be seen in Fig. 1, the final level of responding in the various groups did not fall far short of matching the percentage of reinforcement on the tapes.

None of the main effects or interactions comparing VR alone and VR plus DR were significant. Although varying group size, and consequently proportion of VR, did affect amount of learning in the observer, the lack of additional effect due to DR was not related to group size and the ratio VR to DR, at least within the range examined in the present study.

REFERENCES

ESTES, W. K. Learning. *Annu. Rev. Psychol.*, 1956, **7**, 1–38.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

KANFER, F. H., & MARSTON, A. R. Human reinforcement: Vicarious and direct. *J. exp. Psychol.*, 1963, **65**, 292–296.

LEWIS, D. J., & DUNCAN, C. P. Vicarious experience and partial reinforcement. *J. abnorm. soc. Psychol.*, 1958, **57**, 321–326.

MCBREARTY, J. F., MARSTON, A. R., & KANFER, F. H. Conditioning a verbal operant in a group setting: Direct vs. vicarious reinforcement. *Amer. Psychologist*, 1961, **16**, 425. (Abstract)

(Received August 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 597–602

VISUAL RECOGNITION THRESHOLDS AS A FUNCTION OF VERBAL ABILITY AND WORD FREQUENCY[1]

CHARLES D. SPIELBERGER[2] AND J. PETER DENNY[3]

Duke University

Effects of verbal ability and word frequency on visual-recognition thresholds were investigated. Words of low, moderate, and high frequency of occurrence in the Thorndike-Lorge word counts were tachistoscopically presented to college students selected on the basis of extreme scores on the ACE Linguistic scale. Inverse relationships were obtained between word frequency and recognition thresholds for both high and low verbal ability Ss. High ability Ss recognized low frequency words more rapidly than low ability Ss.

An inverse relationship between the recognition thresholds for words and the frequency of their occurrence in representative samples of written language has been clearly demonstrated in a number of investigations (Howes & Solomon, 1951; McGinnies, Comer, & Lacey, 1952; Postman & Schneider, 1951; Solomon & Howes, 1951; Taylor, Rosenfeldt, & Schulz, 1961). In these studies, the frequency of occurrence of a given word in the Thorndike-Lorge (1944) word counts was assumed to be an index of *S*s' prior usage of the word.

Solomon and Postman (1952) have pointed out, however, that word counts may not be adequate measures of the idiosyncratic frequency of word usage for individual *S*s, and that this potential shortcoming of word counts can be avoided by directly manipulating *S*s' usage of nonsense syllables. Although these investigators and others have convincingly demonstrated that recognition thresholds for nonsense syllables varied inversely with the frequency of prior usage of the syllables (Baker & Feldman, 1956; King-Ellison & Jenkins, 1954; Postman & Rosenzweig, 1956; Taylor, 1958), the effects of the idiosyncratic frequency of usage of *words* on recognition thresholds has apparently been investigated directly in only a single study. Daston (1957) found that individual patients obtained lower thresholds for words which they used with greater frequency in therapeutic interviews than for control words matched with the idiosyncratic words on the basis of frequency of occurrence in the Thorndike-Lorge word counts. In this experiment, however, it was possible that frequency was not the primary determinant of thresholds for idiosyncratic words, since these words were used more recently, as well as more frequently, than control words. Postman and Solomon (1950) found that when frequency was controlled, recency of word usage was significantly correlated with recognition thresholds.

While the relationship between word frequency and recognition thresholds has been subjected to intensive investigation, variables which influence word frequency, and thereby indirectly affect thresholds, have received little attention. For example, verbal ability would appear to be an important determinant of the idiosyncratic usage of words. It is generally accepted that verbal ability is strongly related to the size of an individual's vocabulary (Wechsler, 1958). Indeed, high verbal ability or verbal intelligence is often inferred from vocabulary tests which measure the ability to define correctly the sorts of words which occur with low frequency in word counts. Therefore, it seems reasonable to assume that *S*s with high verbal ability would be more familiar with words which occur with low frequency in word counts than would be *S*s with low verbal ability. There would seem to be little basis for assuming differential frequency of word usage as a function of verbal ability for words which appear in word counts with high frequency.

---

[1] This research was supported in part by a grant (MH-7229) to the first author from the National Institute of Mental Health, United States Public Health Service. We are indebted to David A. Altfeld, Edna Bissette, and Kay Howard for their assistance in the collection and analysis of the data.

[2] Now at Vanderbilt University.

[3] Now at the University of Western Ontario.

The purpose of this experiment was to evaluate the effects of verbal ability and word frequency on visual-recognition thresholds. On the basis of the preceding assumptions, it was hypothesized that verbal ability would interact with word frequency to determine recognition thresholds. More specifically, it was expected that, for words with low frequency of occurrence in the Thorndike-Lorge word counts, thresholds for *S*s of high verbal ability would be lower than for *S*s of low ability; and that, for high frequency words, there would be no difference in recognition thresholds between high and low ability *S*s. Although there was no basis for predicting the effects of verbal ability on recognition thresholds for words which occur in word counts with moderate frequency, such words were also included in this investigation so that the psychophysical relationships between word frequency and recognition thresholds for *S*s who differed in verbal ability could be compared.

METHOD

Subjects.—The *S*s were 32 male students enrolled in the introductory psychology course at Duke University who scored in the upper and lower 20% of the distribution of scores on the Linguistic scale of the American Council on Education Psychological Examination for College Freshmen (ACE), 1949 edition. This test of general scholastic aptitude had been routinely administered as a part of a placement battery to all students entering the university as freshmen. Sixteen *S*s with ACE raw scores of 62 or below were designated the Low Verbal Ability group; a second group of 16 *S*s whose scores were 83 or above on the same scale were designated the High Verbal Ability group. These scores correspond to percentile ranks of 43 and 87, respectively, on the ACE national norms. Students were requested not to volunteer for the experiment if they had any uncorrected visual defects. This was deemed necessary since a significant negative correlation between a measure of visual acuity and recognition thresholds for high frequency words ($r = -.50$) was found in a previous study with college students in which no attempt was made to screen out *S*s with visual defects.

Word stimuli.—It has been demonstrated that word length influences recognition thresholds, and that the effects of word frequency on recognition thresholds are more pronounced for long words than for short words (McGinnies et al., 1952). Therefore, all of the experimental words in the present study contained nine letters. Twelve words were selected, four at each of three levels of word frequency. In order to insure that the words were homogeneous at each frequency level, narrowly defined frequency ranges on *both* the Thorndike-Lorge Semantic (S) and Magazine (L) counts were employed. The criteria for the selection of word stimuli were: for high frequency words, a frequency of 1,000 or greater on one of the two word counts with a frequency of 200 or greater on the other; for moderate frequency words, 100–200 on both word counts; and for low frequency words, 2–10 on both counts. The words also met the following criteria: (*a*) they had three or four syllables; and (*b*) they were not compounds or roots of other words with markedly different frequencies. The stimulus words in the order in which they were presented to each *S* are given in Table 1

TABLE 1

Experimental Words Listed in Order of Presentation to Ss in Exp. II and the Frequency of Their Occurrence in the Thorndike-Lorge S and L Counts

Words	Frequency of Occurrence	
	S Count	L Count
NECESSARY	1,017	715
RECOMMEND	161	154
BARRISTER	6	2
NOCTURNAL	7	5
ASSISTANT	173	151
AFTERNOON	283	1,033
CONDITION	1,775	855
CURIOSITY	155	157
AFFLUENCE	8	6
HIBERNATE	3	5
SENSATION	156	124
IMPORTANT	1,350	1,130

along with the frequency of occurrence of each word in the S and L counts. Six practice words (PAY, AGE, FOR, PRESIDENT, DIFFERENT, SOMETHING), all with frequencies of 1,000 or greater on both word counts, were presented prior to the experimental words.

Procedure.—The *S*s were tested individually. Measures of near and far visual acuity were obtained for each *S* by means of an American Optical Company orthorater, employing checkerboard targets (Department of the Army, 1948). In the measurement of visual acuity and in the determination of word thresholds, *S*s were instructed to use corrective glasses if these were customarily worn.

The *S* was then seated before the viewing window of a Gerbrands mirror tachistoscope in a darkened room and instructed as follows:

> This experiment is a test of your ability to recognize words that are presented very quickly. These are all ordinary words, some more familiar than others. The word will appear in this window in the center of the lighted area. First, each word will appear for such a short period of time that you will not recognize it. However, I will gradually increase the exposure time until you have reported the word correctly on two trials. After you have reported the word correctly twice, I will tell you that you are correct. Report what you see after each exposure. Say the word if you recognize it. If you are not quite sure, you should try to guess the word. If you do not recognize any word, spell out whatever letters you have seen. If you see nothing distinctly, say *none.* Before each presentation of a word, I will say *ready.* At the ready signal, focus your eyes on the rectangular patch of light.

The stimulus words were then presented individually in the same fixed order to each *S* at a distance of approximately 23 in. from *S*'s eyes. In order to equate the effects of practice on low (L), moderate (M), and high (H) frequency words, the order of presentation of these words was counterbalanced in the following manner: H-M-L-L-M-H-H-M-L-L-M-H. Each word was initially presented five times at an exposure time of .01 sec. The duration of exposure was increased in .01-sec. steps with five presentations at each exposure time until *S*'s recognition threshold was reached. The recognition threshold of a word was defined as the total number of exposures required for *S* to reach the criterion of two, not necessarily consecutive, correct reports of the word. Two alternative criteria were also employed for defining thresholds. These were: three reports with six or more letters correct; or two such reports combined with a single correct report of the word. The alternative criteria were used because it was found in pilot studies that an occasional *S* of low verbal ability did not know that some of the low frequency words were actual words.

RESULTS

The mean numbers of exposures required for the recognition of high, moderate, and low frequency words by *S*s with high and low verbal ability are presented in Fig. 1. These data were evaluated by an analysis of variance (Lindquist, 1953) in which Verbal Ability was a between-*S*s variable and Word Frequency a within-*S*s variable. The obtained highly significant main effect of Word Frequency, $F\ (2, 60) = 67.49, p < .001$, reflected in the curves for both high and low verbal ability *S*s in Fig. 1, was generally consistent with the results of previous studies in which substantial inverse relationships between word frequency and recognition thresholds were found. The predicted Word Frequency × Verbal Ability

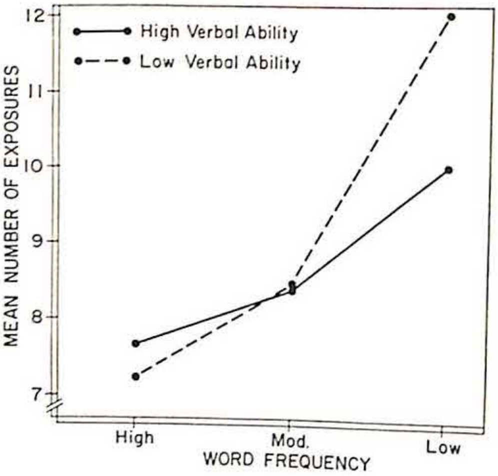


Fig. 1. The mean number of exposures required for recognition of high, moderate, and low frequency words by *S*s of high and low verbal ability.

interaction, $F(2, 60) = 8.34, p < .001$, was also highly significant, indicating that the psychophysical function relating word frequency to recognition thresholds differed for *S*s with high and low verbal ability. The finding that *S*s of high verbal ability required fewer exposures for the recognition of low frequency words than did an *S* of low verbal ability, as may be noted in Fig. 1, was consistent with expectations based upon the assumption that the high ability *S*s were more familiar with infrequent words than low ability *S*s.

It may also be noted in Fig. 1 that *S*s with low verbal ability had slightly lower thresholds for high frequency words. In order to determine if this was related to uncontrolled differences between groups in visual acuity, the orthorater visual acuity scores (far plus near) for *S*s of high and low verbal ability were compared. No differences between these groups in visual acuity were reflected in orthorater scores which were found to be positively, albeit nonsignificantly, correlated with ACE scores ($r = .14$).

Since it has been observed that there are marked effects of practice in visual-recognition threshold experiments (Howes & Solomon, 1951), the effects of practice in this experiment were evaluated by comparing the thresholds for the first six experimental words with those of the last six words. The mean number of exposures required for recognition of words in the first and last half of the experimental word list were 8.90 and 9.03, respectively, and the difference between these means was not found to be statistically significant, $t(30) = 0.79$. Thus, it would appear that the utilization of six practice words succeeded in eliminating practice effects.

Discussion

The finding that word-recognition thresholds were inversely related to the frequency of occurrence of words in the Thorndike-Lorge word counts was generally consistent with the inverse relationships obtained between frequency and thresholds in previous investigations (Howes & Solomon, 1951; McGinnies et al., 1952; Postman & Schneider, 1951; Solomon & Howes, 1951; Taylor et al., 1961). In the present study, however, *S*s of high verbal ability required fewer exposures to recognize low frequency words than did *S*s of low verbal ability. This finding suggested that the high ability *S*s were more familiar with the low frequency words than were the low ability *S*s. The assumption that verbal ability determines idiosyncratic frequency of usage of low frequency words was consistent with but not dependent upon the demonstration of an empirical relationship between verbal ability and recognition thresholds. The results of the present study would seem to suggest that word counts systematically underestimate the frequency of usage of low frequency words for high ability *S*s, and/or overestimate the usage of such words for low ability *S*s.

The interpretation of the findings of the present study in terms of the idiosyncratic frequency of prior word usage by *S*s who differed in verbal ability implicitly assumes that the effects of frequency on recognition thresholds are mediated by verbal habits which determine the probability that a word will be reported under experimental conditions in which minimal stimulus cues are provided. The importance of such habits as determinants of recognition thresholds was clearly demonstrated in a recent experiment in which nonsense syllables were presented with differential frequency in a training period prior to the tachistoscopic exposure of blank stimuli (Goldiamond & Hawkins, 1958). In this study, an inverse relationship between frequency and "recognition thresholds" was found even when no word stimulus was presented. This finding strongly supports the hypothesis that the effects of word frequency on recognition thresholds are mediated by learned response dispositions (Postman, 1953).

Although we have emphasized word frequency in the interpretation of our results, the findings of the present study may also be explained in terms of differences in the level of "perceptual organization" of *S*s with high and low verbal ability. Krech and Calvin (1953) reported a high correlation (biserial $r = .91$) between level of perceptual organization, as measured by *S*s' abilities to reproduce tachistoscopically presented dot patterns, and scores on the Wechsler Vocabulary Test, a measure of verbal intelligence. This finding was corroborated in a second study (Calvin, Tyrrell, Schulz, & Koons, 1958) in which the perceptual task employed by Krech and Calvin was described as similar to the Raven Progressive Matrices Test. According to Raven (1938), a "correct" perception on this test depends on

> a person's capacity at the time of the test to apprehend meaningless figures presented for his observation, see the relations between them, conceive the nature of the figure completing each system of relations presented, and by so doing, develop a systematic method of reasoning (p. 1).

It might be argued that the "correct" perception of a word stimulus presented at subthreshold exposures depends upon similar processes and that, in the present study, the relationship obtained between verbal ability and thresholds was mediated by differences in perceptual organization in *S*s who differed in verbal ability.

When the findings in the present investigation are considered within the context of those obtained by Krech and Calvin (1953) and Calvin et al. (1958), the possibility must be seriously entertained that differences in verbal ability reflect both (*a*) idiosyncratic frequency of prior usage of low frequency words and (*b*) level of perceptual organization, and that these in turn contribute to the determination of recognition thresholds.

REFERENCES

BAKER, K. E., & FELDMAN, H. Threshold luminance for recognition in relation to frequency of prior exposure. *Amer. J. Psychol.*, 1956, **69**, 278–280.

CALVIN, A. D., TYRRELL, S., SCHULZ, R. E., & KOONS, R. G. The effect of exposure-time on the relation between perceptual organization and intelligence. *Amer. J. Psychol.*, 1958, **71**, 573–577.

DASTON, P. G. Perception of idiosyncratically familiar words. *Percept. mot. Skills*, 1957, **7**, 3–6.

DEPARTMENT OF THE ARMY, Personnel Research Branch. Studies in visual acuity. *USA TAGO Personnel Res. Br. Rep.*, 1948, No. 742.

GOLDIAMOND, I., & HAWKINS, W. F. Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 1958, **56**, 457–463.

HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.

KING-ELLISON, P., & JENKINS, J. J. The durational threshold of visual recognition as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 700–703.

KRECH, D., & CALVIN, A. Levels of perceptual organization and cognition. *J. abnorm. soc. Psychol.*, 1953, **48**, 394–400.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

McGinnies, E., Comer, P. B., & Lacey, O. L. Visual-recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44**, 65–69.

Postman, L. The experimental analysis of motivational factors in perception. In, *Current theory and research in motivation: A symposium.* Lincoln: Univer. Nebraska Press, 1953.

Postman, L., & Rosenzweig, M. R. Practice and transfer in the visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 1956, **69**, 209–226.

Postman, L., & Schneider, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271–284.

Postman, L., & Solomon, R. L. Perceptual sensitivity to completed and incompleted tasks. *J. Pers.*, 1950, **18**, 347–357.

Raven, J. C. *Guide to using progressive matrices.* London: Lewis, 1938.

Solomon, R. L., & Howes, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.

Taylor, J. A. Meaning, frequency, and visual duration thresholds. *J. exp. Psychol.*, 1958, **55**, 329–334.

Taylor, J. A., Rosenfeldt, D. C., & Schulz, R. W. The relationship between word frequency and perceptibility with a forced-choice technique. *J. abnorm. soc. Psychol.*, 1961, **62**, 491–496.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

Wechsler, D. *The measurement and appraisal of adult intelligence.* (4th ed.) Baltimore: Williams & Wilkins, 1958.

(Received August 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 603–608

POSITION MEDIATED TRANSFER BETWEEN SERIAL LEARNING AND A SPATIAL DISCRIMINATION TASK

SHELDON M. EBENHOLTZ

Connecticut College

In a position learning task (PL) Ss learned to discriminate between 10 spatial locations arranged in a vertical array, by associating a nonsense syllable with each position. Transfer from serial learning (SL) to PL and from PL to SL was studied. In Condition C the temporal order of syllables in SL (i.e., from beginning to end) corresponded with the spatial location of the same syllables in PL (i.e., from top to bottom). The relative positions of the syllables was random in Condition D. In Condition R the relative syllable positions were displaced, however adjacent items in SL were at contiguous locations in PL. Transfer in both directions was highest in the 1st condition. With one exception, the results of Conditions D and R did not differ from each other. Results yield evidence for position learning in SL and raise the question of the role of sequential association.

Recent attempts to isolate the effective stimulus in serial learning (Ebenholtz, 1961, 1963; Horowitz & Izawa, 1963; Young, 1962) suggest the operation of two variables. These may be referred to as sequential association and temporal position. Although it is most probable that the formation of specific associations between successive terms is a sufficient condition for mastery of a serial list, present evidence does not warrant the same statement with regard to knowledge of the positions of the component items.

The present experiment sought to isolate the role of position and to test the effectiveness of knowledge of position as a mediator of serial learning.

METHOD

A situation was sought which might represent a pure case of position learning; viz., one with minimum possibility of forming sequential associations. A spatial position learning task was devised with this purpose in mind.

The *S*s were trained to discriminate between a vertical array of 10 small rectangular "windows" such that given any window *S* could supply the syllable appropriate to it. On the assumption that serial learning entails the ordering of syllables with respect to temporal location, both tasks involve a common process; i.e., position learning. Transfer from one task to the other should occur readily providing the syllables are ordered in similar fashion; viz., providing Syllable 1 in serial learning has to be associated with the top position in position learning, etc. If serial learning does *not* entail position learning, then the relative placement of syllables across the tasks should not effect the degree of transfer. A control condition, therefore, was introduced in which the relation of the positions of syllables across the tasks was random; e.g., the first syllable in serial learning appeared toward the center of the spatial array, etc. A second control was introduced in which the items maintained their relative order on the two tasks but were displaced from their absolute positions. This condition permitted the evaluation of the extent to which transfer may be mediated by specific associations between adjacent items.

Procedure

Position learning (PL).—The *S* was seated in front of a column of $10\frac{1}{4} \times \frac{1}{2}$ in. rectangular windows cut in a large white cardboard screen. There was an interspace of $\frac{3}{4}$ in. between windows. Behind the display was a large revolving drum 14 in. high and 18 in. in diameter. Nonsense syllables were placed on the drum so as to appear in predetermined order in one window at a time. The apparatus functioned much like a traditional memory drum turned on its side with the long axis in vertical position.[1]

The purpose of the position learning task was to have *S* respond on the basis of spatial location. This condition may, therefore, be regarded as a paired-associate task with window location as the stimulus and nonsense syllable as response. In order to indicate the window to which *S* was to respond, a red patch was presented, all other windows remaining blank. After a fixed interval the red patch disappeared and the syllable appropriate to that position took its place.

The temporal order of presentation of syllables was determined by five orders of a 10-item Latin square. Items which occurred at spatially contiguous locations were never presented in temporal succession. After five trials the first five orders of presentation were repeated.

The *S* was allowed 3 sec. to respond with the prescribed syllable while the red patch was in view. The syllable itself appeared for an additional 3 sec. There was a 12-sec. intertrial interval. The *S*s learned to a criterion of one perfect anticipation of all 10 items.

Serial learning (SL).—The conventional method of serial learning was used. However, conditions were made as similar to those of PL as possible. A red patch appeared prior to the first item and between each of the succeeding items. The *S*s were instructed to respond with the succeeding term while the red patch was in view. Both the patch and the item were presented for 3 sec. each. The criterion, presentation rate, and intertrial interval were identical in PL and SL. In both conditions *S* was requested to pronounce the syllables and to correct himself in the event of a wrong response or no response. The syllables in both PL and SL were from the Glaze (1928) list of 80–100% association value. Intralist similarity was minimal.

Design

A transfer design was employed to study the effects of PL on SL and of SL on PL. Two groups of *S*s learned in the order of SL-PL. Under Cond. C the syllables in SL and PL were maintained at *coordinate* positions in both tasks. That is Syllables 1 through 10 in SL appeared at Windows 1 through 10, respectively, in PL.[2] However, the temporal order of presentation of syllables in PL was systematically varied and never corresponded with the temporal order of syllables in SL. In short, the temporal position of syllables in SL corresponded with their spatial position in PL.

Under Cond. D the items were presented at *disparate* locations across the tasks. In this condition Syllables 1 through 10 in SL appeared in PL at Windows 8, 6, 4, 10, 1, 9, 7, 2, 5, and 3, respectively. Two additional groups learned in the order PL-SL. One group learned under Cond. C and the other learned under Cond. D. There was a 2-min. interval between tasks. In order to facilitate comparisons between Cond. C and D it was necessary to counterbalance lists across these conditions. For this purpose two forms of an SL and PL list were used. The relations between the two SL and the two PL lists are summarized in the following paradigm:

$$\begin{tikzcd} SL_1 \arrow[r, dash] \arrow[dr, dash] & PL_1 \arrow[dl, dash] \\ SL_2 \arrow[r, dash] & PL_2 \end{tikzcd}$$

The subscripts refer to the different forms of the lists used in SL and PL, respectively. When SL and PL have common subscripts Cond. C is indicated (viz., SL_1-PL_1 and SL_2-PL_2); where subscripts differ the relation between items is that of Cond. D. The net result of the counterbalancing procedure was that the first learned lists of Cond. C and D, respectively, were identical as were the lists learned secondly. Two additional groups of *S*s constituted a third condition (R) in which the *relative* order of items on SL and PL was preserved. These were run after Cond. C and D had been completed. One group learned in the order SL-PL, the other in the order PL-SL. The PL list was identical with one of the forms used in Cond. C and D. The present SL list contained the same terms as those used previously, however, the presentation sequence differed. Items 1 through 10 occurred at Windows 8, 9, 10, 1, 2, 3, 4, 5, 6, and 7, respectively, in PL. Thus except for the sequence between Items 3 and 4 in SL all terms which appeared in temporal succession in SL now appeared in spatially contiguous locations in PL. It may be noted that Cond. R and C were similar in that the *relative* order of terms was maintained across tasks. They differed, however, in that under Cond. R items occurred at disparate positions

---

[1] The author wishes to thank Otto Hebel for his aid in the design and construction of the apparatus.

[2] The numbers represent the order of windows from top to bottom.

TABLE 1

Mean Trials to Criterion and Errors per Trial on Initial List

Cond.	Mean Trials to Criterion				Mean Errors Per Trial			
	SL		PL		SL		PL	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C	7.00	2.36	9.40	2.87	3.94	1.28	4.63	.98
D	7.10	2.84	9.80	2.44	3.20	1.02	4.53	1.00
R	7.50	2.94	10.42	2.17	3.53	.75	4.67	.60

on SL and PL, respectively, e.g., Item 5 in SL occurred at Position 2 in PL, etc. In this latter respect Cond. R was similar to Cond. D.

The purpose of Cond. R was to check on the possibility of the transfer of specific associations between the temporally adjacent items of SL and the spatially contiguous terms of PL. If such associations may be assumed to occur under Cond. C (e.g., by rehearsal of groups of items in sequence) then high transfer to SL-C and low transfer (or interference) to SL-D is to be expected. The same possibility holds in the case of transfer from SL to PL.

Since position mediated transfer is excluded under Cond. R and D, the present view requires maximum transfer to occur under Cond. C. If performance under Cond. R and D is equivalent then the role of specific association must be assumed to be negligible.

Subjects

A total of 40 Connecticut College students took part in Cond. C and D and 28 others in Cond. R. None had previous experience with rote learning in the laboratory. Of the former group 20 *S*s had learned in the order SL-PL and 20 in the order PL-SL. Ten *S*s in each of these groups learned under Cond. C and D, respectively, with half of the 10 *S*s learning either Form 1 or Form 2 of their SL or PL lists. The *S*s were assigned to Cond. C and D alternately, as they appeared at the laboratory. Those *S*s taking part in Cond. R were assigned alternately to task order SL-PL or PL-SL.

RESULTS

The results of Cond. C and D are based upon the combined data of the two forms of the SL and PL lists.

First list.—On the basis of t tests all comparisons within SL and PL, respectively, between Cond. C, D, and R revealed no significant differences in original learning. Inspection of Table 1 indicates, however, that the PL task was apparently more difficult than SL.

Evaluation was accomplished by pooling the data of all three groups on PL and SL, respectively. For trials to criterion $t\ (66) = 4.23, p < .01$. Comparison of errors per trial yielded $t\ (66) = 8.15, p < .01$.

TABLE 2

Mean Trials to Criterion and Errors per Trial on Transfer List

Cond.	Mean Trials to Criterion				Mean Errors Per Trial			
	SL		PL		SL		PL	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C	1.60	.48	5.10	2.38	.40	.37	2.22	1.08
D	4.20	.97	10.00	2.60	2.66	1.01	4.19	.60
R	3.35	1.99	8.42	2.02	1.44	1.01	3.73	.59

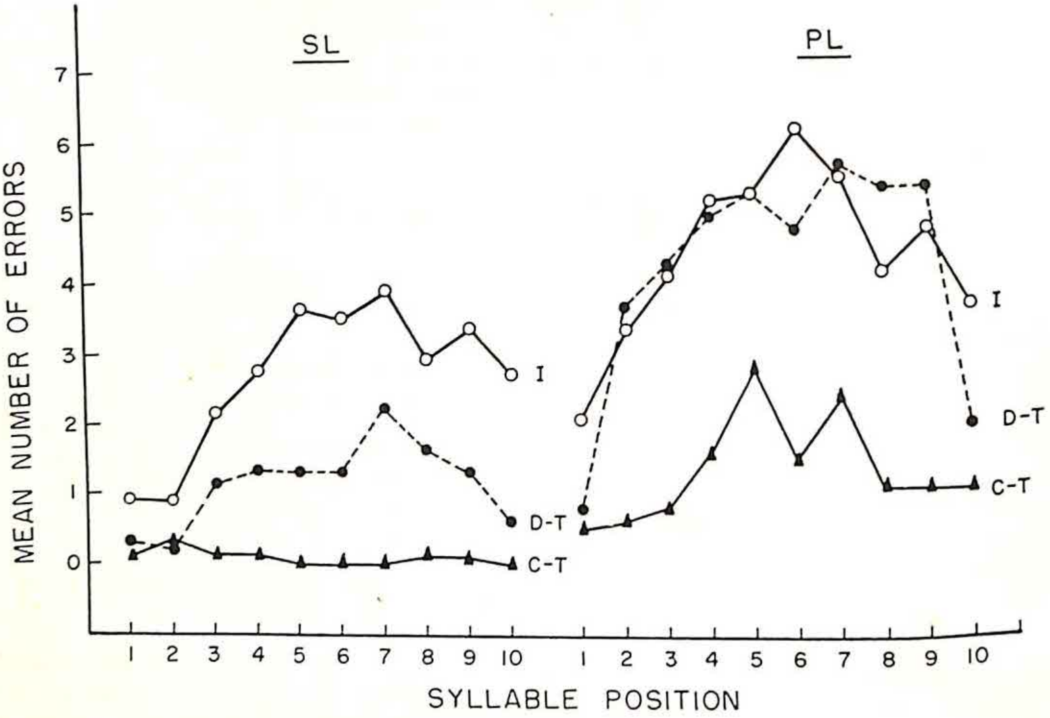


FIG. 1. Mean errors per position under SL and PL, Cond. C and D, respectively, and initial (I) and transfer (T) tasks.

Second list (transfer).—The results of the transfer task are represented in Table 2. When SL is the transfer task, Cond. C is learned with significantly fewer trials to criterion and less errors per trial than Cond. D; $t\ (18) = 7.22$ and 6.64, in each case $p < .01$. Similarly performance under SL-C was superior to SL-R, comparisons yielding $t\ (22) = 2.36$, $p < .05$ and $t\ (22) = 3.71$, $p < .01$ for trials and errors per trial, respectively. Condition D did not differ significantly from Cond. R in trials to criterion, however, the latter did produce significantly fewer errors per trial; $t\ (22) = 2.97$, $p < .01$.

Considering PL as transfer task, Cond. D and R did not differ significantly in either trials to criterion or errors per trial. On both indices Cond. C proved favorable in comparison with both Cond. D and R. Comparing PL-C with PL-D, $t\ (18) = 4.19$, $p < .01$, and $t\ (18) = 4.92$, $p < .01$, for trials to criterion and errors per trial, respectively. Condition PL-C required significantly fewer trials, $t\ (22) = 3.45$, $p < .01$ and less errors per trial, $t\ (22) = 4.08$, $p < .01$, than Cond. PL-R.

It is of some interest to note that the mean scores of Cond. R consistently fall between Cond. C and D (see Table 2). The same is true in comparisons of percent transfer. This may be represented by comparing performance on the transfer list (T) with performance on the same task as initial list (I). The formula $(I-T)/I \times 100$ with trials to criterion as index, yielded 77.14, 55.33, and 40.84, for Cond. SL-C, SL-R, and SL-D, respectively, and 45.74, 19.19, and -2.04 for PL-C, PL-R, and PL-D in that order. These results indicate more transfer to SL than to PL under each condition, with the conditions arranging themselves in the order C-R-D.

Position effects.—Both SL and PL lend themselves to analysis of errors by position. This is shown in Fig. 1. In order to plot the data, the top window in PL was arbitrarily designated as No. 1, the second from the top as No. 2, etc.

The curves of the initial SL and PL lists are based upon the combined data of Cond. C and D. The SL and PL curves are similar in that under both conditions end items are learned rapidly, central items yielding maximum errors. Such similar distributions are to be expected if identical processes (i.e., position learning) underlie both tasks.

The large amount of transfer under SL-C virtually eliminated the serial position curve. Nine of the 10 positions in SL yielded fewer errors when transfer occurred under Cond. C as compared with Cond. D. Condition PL-C yielded fewer errors than PL-D at all positions. Although not plotted, the error data of Cond. R fell between the curves of Cond. C and D. This was true of both SL and PL.

DISCUSSION

In comparing Cond. C and D it was demonstrated that prior SL facilitated subsequent PL and vice versa, and that this occurred more readily when items appeared at coordinate positions on both tasks (Cond. C) rather than at disparate locations (Cond. D). This result suggests that knowledge of position is a significant factor in serial learning. This conclusion is valid only to the extent to which the operations defining PL exclude the forming of associations between spatially contiguous items. With one exception (SL-R) the results of Cond. R indicated that the preservation of the relation of contiguity between adjacent items in SL and PL, respectively, produced transfer which was not significantly greater than when the temporally contiguous terms of SL appeared at nonadjacent locations in PL (Cond. D). Since Cond. R produced little or no transfer beyond that of Cond. D, it follows that the facilitative effects of Cond. C were mediated primarily by the common positions of the items across the tasks.

The fact that transfer from PL to SL exceeded that from SL to PL raises a question. If both tasks entailed the learning of position, then transfer should have been equal in both directions. One explanation is that while both tasks involved position learning to a common criterion they differed in the degree to which the positions were discriminable at the time of transfer. Differences in position discriminability may also have contributed to initial differences in task difficulty. The fact that the initial tasks were of unequal difficulty, thereby producing differences in practice and item familiarity, must also be considered as possible explanations. Another question relates to the fact that transfer was less than perfect. This may have been due to a number of variables. Some loss may be attributable to the shift from the spatial to temporal modalities, and to the small time interval between tasks. In addition it is possible that the degree of learning of the initial tasks may have been inadequate for maximum transfer. A more important consideration, however, is that position learning may *not* be adequate to account *completely* for serial learning. To the extent to which specific associations are required, the present conditions cannot yield complete transfer. The fact that Cond. R consistently produced transfer data which fell midway between Cond. C and D is support for the assumption that to some extent sequential associations are formed in serial learning.

It is of some interest to compare the effects of transfer obtained by Young (1959) with those of the present study. Young's *S*s first learned paired associates and then transferred to a serial list in which pairs of successive terms were identical with those of the paired-associate list. If specific associative bonds were sufficient to master the serial list, then ideally Young's *S*s should have learned within one trial. Actually a mean of 8.08 trials was required to reach one correct recitation of the entire list. Since in the present study transfer to SL-C was likewise less than perfect, it appears reasonable to conclude that neither position nor association are sufficient to explain serial learning. It seems most likely that both processes play a role.

The question remains as to the actual utilization of each process in the course of learning a serial list. Available evidence has not clarified this problem. There is some evidence (Ebenholtz, 1961) to suggest that responding on the basis of an association is more difficult than responding to position. On this assumption one may conceive of serial learning as proceeding on the basis of temporal position (i.e., at the ends of the list) up to the point where position discrimination is too poor to produce additional correct anticipations (i.e., at the center of the list). The *S*s must then form consecutive associations between the terms at these poorly differentiated positions. The bow-shaped serial position curve may be understood as a consequence of this analysis by assuming position discrimination to be most difficult at the center of the list (see Schulz, 1955, for corroborative evidence). Error analysis of the spatial position task (see Fig. 1 and also Jensen, 1962) supports this view, as well.

It is important to note that the present formulation is contradictory to that offered by Young (1962) that "specificity learning occurs at the extremes and that position learning occurs in the middle of the list" (p. 313). Thus there is need for further clarification although there is good agreement as concerns a dual process theory of serial learning.

References

Ebenholtz, S. M. The relative roles of position learning and sequential association in the serial learning process. Unpublished doctoral dissertation, New School for Social Research, 1961.

Ebenholtz, S. M. Serial learning: Position learning and sequential associations. *J. exp. Psychol.*, 1963, in press.

Glaze, J. A. The association values of nonsense syllables. *J. genet. Psychol.*, 1928, **25**, 255–269.

Horowitz, L. M., & Izawa, C. Comparison of serial and paired-associate learning. *J. exp. Psychol.*, 1963, **65**, 351–362.

Jensen, A. R. Temporal and spatial effects of serial position. *Amer. J. Psychol.*, 1962, **75**, 390–400.

Schulz, R. W. Generalization of serial position in rote serial learning. *J. exp. Psychol.*, 1955, **49**, 267–272.

Young, R. K. A comparison of two methods of learning serial associations. *Amer. J. Psychol.*, 1959, **72**, 554–559.

Young, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *J. exp. Psychol.*, 1962, **63**, 307–313.

(Received August 6, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 609–615

SERIAL VERSUS RANDOM PRESENTATION OF PAIRED ASSOCIATES[1]

CLESSEN J. MARTIN AND ELI SALTZ

Center for the Study of Cognitive Processes, Wayne State University

The effect of constant serial order in presentation of paired associates was investigated in 2 experiments. Experiment I (100 Ss) found that constant order resulted in serial associations between the response items of the pairs (significant beyond .01) but, contrary to other studies, did not facilitate paired-associate learning. Experiment II (72 Ss) tested the hypothesis that facilitory effects due to constant order were a function of S-R presentation rate. Contrary to prediction, no facilitory effect occurred at any presentation rate. Results indicate that, contrary to previous belief, serial presentation of paired associates does not necessarily facilitate learning.

Studies by McGeoch and McKinney (1937) and by McGeoch and Underwood (1943) have reported that presentation of S-R pairs in the same serial order on each learning trial will facilitate learning. However, it was not possible to determine the exact source of such facilitation in these studies. Newman and Saltz (1962) have attempted to identify more clearly the source of serial facilitation in paired-associate learning.

To test the hypothesis that facilitation is produced by serial position cues, these investigators, following the procedure of McGeoch and McKinney, employed a condition in which S-R pairs were presented in a constant serial order for five trials, then on the sixth trial *S*s were tested with the stimuli presented in random order. Contrary to the hypothesis that facilitation is produced by serial cues, this group performed no worse on the sixth trial than a group in which pairs were presented in a constant serial order for the five learning trials and also the sixth, *test*, trial. Both serial groups performed better than a random presentation group on the sixth trial.

The purpose of Exp. I in the present paper is to investigate the possibility that the lack of difference between the serial presentation-serial test, S(S) and the serial presentation-random test, S(R), groups in the Newman and Saltz study was due to the fact that their test situation was not sufficiently sensitive to permit an adequate test of the serial position cue hypothesis. Since the degree of learning over five trials was low due to competing S terms in the Newman and Saltz investigation, the S terms in the present study were noncompeting and of higher association value in order to increase the degree of learning over the first five learning trials. In addition, after the test on Trial 6, an additional five learning trials and a final test trial were administered to *S*s to determine if the Newman and Saltz results held up over greater numbers of trials. A 10-sec. interval was used for each presentation of an S term during the retention test in the Newman and Saltz investigation. A 2-sec. interval was used in the present study.

---

[1] The present study is part of a program of research, concerned with the influence of differentiation on verbal processes, supported by the National Science Foundation under NSF G-9068 to the junior author.

It was thought that the 10-sec. interval for each presentation of the S terms reduced the sensitivity of retention measurement and that there may have been differences between the S(S) and S(R) groups which failed to emerge because of the reduced sensitivity of the 10-sec. interval.

The procedure, employed by Newman and Saltz, of presenting the S-R pairs for five trials before the first test trial, was retained in the present study. This procedure was considered critical for a test of the hypothesis that serial position cues were responsible for the facilitation of the constant order conditions. The crucial aspect of the present study is the performance of the group trained on a constant order of presentation when shifted to a random order. The results of the McGeoch and McKinney and the McGeoch and Underwood studies suggest that the use of the standard anticipation method during the first five training trials should lead to a history of more successful anticipations during these training trials for the constant order group than for the random presentation group. When the serial position cues are eliminated for the constant order group by randomizing the order on Trial 6, a failure of this group to fall to the level of the random presentation group might be attributable to the greater history of successes during learning trials.

Finally, to determine if serial presentation of pairs resulted in learning the serial order of responses, after paired-associate learning, *S*s learned a serial list composed of the R terms of the original S-R pairs.

EXPERIMENT I

Method

Subjects.—The *S*s were 100 introductory psychology students at Wayne State University. The first 80 of these were assigned randomly to the experimental conditions with the limitation that subgroup *N*s be equal. After the data had been analyzed a new set of materials was developed and the additional 20 *S*s were randomly assigned to an S(S) or an R(R) treatment involving these materials.

Materials.—List 1, the 10 S-R pairs learned by the first set of 80 *S*s, consisted of the same 10 R terms used in the Newman and Saltz study, high *M* value nouns from Noble's (1952) list; the S terms were nonsense syllables having a 47% association value (Hilgard, 1951). The actual pairs were: YAT-JEWEL, TIS-DINNER, ZUG-MONEY, KEM-VILLAGE, SOZ-INSECT, QUN-GARMENT, XOW-HEAVEN, NOL-WAGON, RUH-OFFICE, BEK-KITCHEN. Two different random orders, an A and a B order, of the above pairs were employed such that half the *S*s in a serial learning condition learned Order A, half learned Order B. An additional five random orders were developed for the random presentation condition learning trials. For all 80 *S*s, the A or B orders were used during test trials. Thus half the *S*s in a random presentation group were tested on Order A, the other half on Order B.

List 2, the materials developed for the 20 additional *S*s, consisted of the identical R terms as for the 80 *S*s, but a new set of 47% association stimuli: FAZ-JEWEL, KOR-DINNER, WAH-MONEY, YOB-VILLAGE, BEW-INSECT, RUV-GARMENT, NAC-HEAVEN, MEP-WAGON, QIG-OFFICE, TAJ-KITCHEN. Material was presented on a Lafayette memory drum.

Procedure.—All *S*s received 10 learning trials in which the stimulus and response members of each pair appeared simultaneously in the window of the memory drum at a 2-sec. rate with 6 sec. between trials. Test trials were presented after the fifth and tenth learning trials. The separation of learning and test trials is consistent with the procedure used by both McGeoch and McKinney (1937) and Newman and Saltz (1962). A test trial consisted of presenting each stimulus member alone for 2 sec. and requiring the *S*s to emit the appropriate response to each stimulus within the 2-sec. interval. In the instructions it was suggested that *S* not attend to the order in which the S-R appeared during the learning trials.

There were four experimental groups.

1. Group S(S)S(S) received the S-R pairs in the same serial order on all learning and test trials.

2. Group R(R)R(R) received the S-R pairs in a different order on each of the first five learning trials and the test trial (i.e.,

six different orders). This sequence of six random orders was then repeated for the second five learning trials and the second test trial.

3. Group S(S)S(R) received the S-R pairs in the same serial order during all 10 learning trials and Retention Test 1, but during Retention Test 2 (twelfth trial) the S terms were presented in a different random order.

4. Group S(R)R(R) received the S-R pairs in the same serial order during the first five learning trials but during Retention Test 1 the S terms were presented in a different random order. In the second five learning trials the S-R pairs were presented in different serial order on each trial. The S terms during Retention Test 2 were presented in the same order as during Retention Test 1.

5. Group S(S) consisted of 10 of the additional 20 *S*s. This group learned List 2. Procedure was identical with that of Group S(S)S(S) except that the *S*s received only five learning trials and one test trial.

6. Group R(R) consisted of the remaining 10 of the additional 20 *S*s. This group also learned List 2. Procedure was identical with that of Group R(R)R(R) except that *S*s received only five learning trials and one test trial.

For Groups 1, 3, and 4, following the second retention test the *S*s were presented with a serial list consisting of the 10 R terms that had been presented during the paired-associate learning trials. Half of the *S*s in each experimental group received the R terms in the identical order in which they had been presented during the learning trials, and the other half received the R terms in an order different from any used during the learning trials. Each *S* was given five trials on the serial list. Each word in the serial list was presented at a 2-sec. rate with a 6-sec. interval between trials. The number of correct anticipations was recorded on each of the last four trials.

Results

Table 1 reveals that serial position did not facilitate the learning of the S-R pairs. The R(R)R(R) group learned more S-R associations than did the S(S)S(S) group as measured by either Retention Test 1 or 2. Additional evidence concerning the effects of serial presentation is provided by the S(S)S(R) group which, through Retention Test 1, also is a serial presentation condition. As can be seen in Table 1, on Retention Test 1 the means for the S(S)S(R) group are close to those of the S(S)S(S) group, and indicate poorer performance than the random group, R(R)R(R). These results are not in accord with those found by previous investigators.

In order to determine whether the difference between the serial and

TABLE 1

MEANS AND *SD*s OF NUMBER OF CORRECT RESPONSES FOR EXP. I ON RETENTION TESTS 1 AND 2

Groups	Order	Retention Test 1			Retention Test 2		
		M	*SD*	Group *M*	*M*	*SD*	Group *M*
1. S(S)S(S)	A	4.4	2.22	4.55	7.3	2.26	7.50
	B	4.7	2.75		7.7	2.59	
2. R(R)R(R)	A	5.5	2.27	5.80	7.5	1.90	7.90
	B	6.1	1.52		8.3	1.34	
3. S(S)S(R)	A	4.9	2.08	4.95	6.9	2.47	7.10
	B	5.0	3.19		7.3	3.20	
4. S(R)R(R)	A	4.2	2.90	4.60	7.0	2.45	7.55
	B	5.0	1.95		8.1	1.45	
5. S(S)	List 2	—	2.26	5.7			
6. R(R)	List 2	—	2.38	6.1			

TABLE 2

Means and *SD*s of Correct Responses in Serial Learning of R Terms of Previously Learned Paired-Associate Lists

Groups and Orders in Original Paired-Associate Learning		Relationship between Serial Order of R Terms in Paired-Associate and Serial Learning					
		Identical			Different		
		M	*SD*	Group *M*	*M*	*SD*	Group *M*
S(S)S(S)	A	16.0	7.07	16.6	11.4	5.59	10.5
	B	17.2	4.32		9.6	4.16	
S(S)S(R)	A	18.6	1.14	18.2	10.2	8.35	10.0
	B	17.8	3.27		9.8	6.22	
S(R)R(R)	A	15.8	6.98	16.4	5.8	4.55	8.6
	B	17.0	7.07		11.4	6.80	

random groups in Exp. I was significant a *t* test was performed. The mean number of correct responses made on Retention Test 1 for the serial groups, Groups 1 and 3, which received the S-R pairs in the same serial position on the first five learning trials was compared to the number of correct responses made by the random group, Group 2, which received the S-R pairs in a different random order on each of the first five learning trials. The mean for the serial groups was 4.75 compared to a mean of 5.80 for the random group. The resulting $t\ (58) = 1.65$ was not significant at the .05 level. A similar comparison of Groups 1 vs. 2 on Retention Test 2 also failed to prove significant.

Groups 5 and 6 were run on a new list, List 2, to determine if the effects reported above were due to idiosyncracies of the S terms employed. (It will be recalled that R terms were those used by Newman and Saltz who found facilitation as a function of serial order.) The findings for Groups 5 and 6 are consistent with the findings for List 1.

With the combined data from Lists 1 and 2, there was a total of 50 *S*s who had received the S-R pairs in a constant serial order during the first five learning trials and had received the S terms in the same serial order on Retention Test 1. There was a total of 30 *S*s for whom each S-R pair appeared in a different serial position during each of the first five learning trials, and during Retention Test 1 the S terms were presented in an order different from any used during the learning trials. The mean number of correct responses on Retention Test 1 for all the R(R) groups was 5.90 as compared to 4.93 for all the S(S) groups. The resulting $t\ (78) = 1.80$ was not significant at the .05 level.

The second phase of Exp. I investigated whether constant order of paired associates leads to serial associations between the R terms of the pairs. Consequently, only *S*s in Groups 1, 3, and 4 were involved in this phase, since these were the only groups in which constant order of pairs was utilized during the earlier, paired-associate phase. Following paired-associate training, *S*s in these three groups were presented a serial list of the 10 R terms they had learned in original paired-associate training. Half the *S*s in each group learned the items in the same serial order in which the items had appeared during paired-associate training. The remaining *S*s learned the items in a new random order. The *S*s received five trials on the serial list, with instructions to anticipate on the last four of these trials. Table 2 gives the means and *SD*s for the correct anticipations made by the various groups on the four anticipation trials.

A 3 × 2 analysis of variance indicates that *S*s receiving the R terms in an identical order to that used during the paired-associate learning task made significantly ($p = .001$) more correct R term anticipations on the serial learning task than did *S*s who received the R terms in an order different from any used during the paired-associate learning task. None of the other effects approached significance.

Originally, the S(S)S(R) and S(R)R(R) groups in Exp. I were run because it was thought that the comparison of these groups with the S(S)S(S) group would give an indication as to the mechanisms involved in facilitation of serial presentation of S-R pairs. However, since no such facilitating effect occurred in Exp. I, the results of the S(S)S(R) and S(R)R(R) Groups of Exp. I become of minor interest.

EXPERIMENT II

The results of Exp. I are contradictory to the findings of McGeoch and McKinney (1937), McGeoch and Underwood (1943), and Newman and Saltz (1962). Presentation of the S-R pairs in the same serial position at every repetition did not lead to a greater degree of learning than presentation in a different serial position each time as measured by Retention Tests 1 and 2. In fact, the group which received the S-R pairs in a changing order learned more S-R associations as measured by Retention Tests 1 and 2 than the group which received the S-R pairs in an unchanging order. Although there was no significant difference between the groups in the mean number of correct responses on Retention Tests 1 or 2, the results were certainly not in the expected direction and are not in accord with other findings. However, the studies reporting facilitation due to serial presentation of S-R pairs involved either easier materials than the present study (McGeoch & Underwood, 1943) or a longer time for recall during the test trials (McGeoch & McKinney, 1937; Newman & Saltz, 1962). Experiment II was designed to determine if a variable related to facilitating learning in general would also act to differentially facilitate learning of S-R pairs in a constant serial order over a random order. The factor chosen for manipulation was rate of presentation, since this permitted using the same paired-associate lists in easy and difficult conditions.

Method

Subjects.—The *S*s were 72 students in introductory psychology at Wayne. The *S*s were randomly assigned to conditions with the limitation that all subgroups have equal *N*s.

Materials.—The three basic lists of Exp. I were used.

Procedure.—A 2 × 3 × 3 factorial design was employed with two rate presentations (a 1-sec. rate and a 4-sec. rate), three sets of materials, and three presentation conditions: S(S), S(R), and R(R), as described in Exp. I.

All *S*s received five learning trials in which the stimulus and response terms were presented simultaneously, followed by a test trial in which only stimulus terms were presented.

Results

Table 3 presents the mean number of correct responses on the retention

TABLE 3

Mean Number of Correct Responses on the Retention Test for Serial and Random Presentation Groups Learning at 1- and 4-Sec. Rates

Groups	1-Sec. Rate		4-Sec. Rate	
	M	*SD*	*M*	*SD*
S(S)	2.83	2.25	8.25	2.05
R(R)	2.25	1.24	7.75	2.34
S(R)	1.67	1.43	7.75	1.87

test (Trial 6) for each group. Each mean is based on 12 *S*s. A 2 × 3 × 3 analysis of variance indicates a significant difference in the number of correct responses between the 1- and 4-sec. time intervals, F (1,54) = 147.44 but there is not a significant difference between the S(S) and R(R) groups nor between lists. The interactions did not approach significance. A t test was used in order to determine whether the S(S) 1-sec. group, was significantly different from the R(R) 1-sec. group. The resulting t (46) = .79 was not significant. A t test between the S(S) 4-sec. group, and the R(R) 4-sec. group, did not yield a significant t value. Thus, although the mean number of correct responses in the S(S) groups was greater than the mean number of correct responses in the R(R) groups, there was no evidence which would support a rejection of the null hypothesis between the means of the S(S) and R(R) groups at either the 1- or 4-sec. interval.

Examining the comparable data relevant to serial vs. random presentation obtained in Exp. I and II yields a relationship of serial to random presentation across three rates of presentation. The 1-sec. and 4-sec. groups of Exp. II each consisted of one third of the *S*s on List 1, Order A; one third on List 1, Order B; and one third on List 2. To provide a similar distribution of materials from Exp. I, Groups 1 and 5 of that experiment were used to determine the mean for serial presentation at 2 sec., and Groups 2 and 6 were used to determine the mean for random presentation at 2 sec. A 2 × 3 × 3 factorial design was used to analyze the results. Again a highly significant effect due to rate of presentation was found. However, there was no significant effect due to serial vs. random presentation ($F < 1.00$) nor did any of the interactions approach significance.

Discussion

Experiment I in the present paper clearly shows that presentation of pairs in a constant order led to formation of serial association between the response items of the pairs. Clearly these serial associations must have been present during paired-associate learning; nevertheless, they did not facilitate that learning. Newman and Saltz (1962) found that constant order of pairs during five learning trials facilitated performance on the test trial (Trial 6) even when pairs were presented in a random order during the test trial eliminating serial position cues. Taken together, the present studies and the Newman and Saltz study indicate that serial position cues are not an important factor in learning paired associates in a constant order.

An unexpected finding arose in connection with the R term serial learning. In the S(R)R(R) condition, the block of five serially presented pairs was followed by five trials in which the R terms appeared in five different orders. There is no evidence, in Table 3, that this random presentation interfered with the subsequent serial learning of the R terms.

The most important conclusion from the present studies is that, contrary to previous belief, it is not necessarily true that serial presentation of S-R pairs will facilitate learning. Nor is it clear under what conditions such facilitation will occur.

REFERENCES

HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Ch. 15.

MCGEOCH, J. A., & MCKINNEY, F. Studies in retroactive inhibition: The influence of the relative order of presentation of original and interpolated paired associates. *J. exp. Psychol.*, 1937, **20**, 60–83.

MCGEOCH, J. A., & UNDERWOOD, B. J. Tests of the two-factor theory of retroactive inhibition. *J. exp. Psychol.*, 1943, **32**, 1–16.

NEWMAN, S. E., & SALTZ, E. Serial position as a cue in learning. *Amer. J. Psychol.*, 1962, **75**, 101–108.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

(Received August 8, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 616–617

SUPPLEMENTARY REPORTS

PRESENTATION TIME, TOTAL TIME, AND MEDIATION IN PAIRED-ASSOCIATE LEARNING: SELF-PACING[1]

B. R. BUGELSKI AND J. RICKWOOD
University of Buffalo

The present report supplements an earlier study where it was shown that while Ss who have long presentation times learn paired associates in fewer trials than Ss with short presentation times, total learning time is the same. 20 college student Ss controlled their own presentation times for the same material. Their total learning times did not vary significantly from the earlier study. It is concluded that a fixed amount of time is required to learn a fixed amount of material regardless of distribution of time over continuous trials.

In the earlier study reported by Bugelski (1962) it was concluded that Presentation Time × Trials (or "total" time) is constant for learning paired-associate nonsense syllables. The conclusion was based on the finding that different *S*s, given the opportunity to examine the response item of a pair of syllables for 2, 4, 6, 8, or 15 sec., took fewer trials to learn eight pairs of syllables as the presentation time increased but spent about the same total amount of time in the effort. It was possible that some of the *S*s were wasting time on the longer presentation trials and that in some manner some artifact might have been introduced. To eliminate this problem, the present study was designed to permit *S*s to pace themselves and terminate the presentation time whenever they desired. In this way, the total time scores would not be artificially extended for the *S*s needing only a few trials to learn.

Method.—The Hunter Card Master was again employed with *S* controlling the manual switch by which the response item could be removed. The next stimulus syllable would then appear after a 2-sec. delay. The stimulus syllable would appear alone for the next 2 sec. during which *S* could anticipate the response syllable. The response syllable would then appear and remain until *S* eliminated it. Thus, while *S* did not control the entire sequence of times, he did have charge of the response time which was systematically varied in the prior experiment. The *S*s were 20 college sophomore volunteers from an experimental psychology class.

Results and discussion.—Although the generalization drawn from the prior study was not necessarily at stake in the present situation it was expected that this generalization would be strengthened if the present results fell within the limits of the total times previously established. The mean total time (see Table 1) for the 20 *S*s (computed per pair learned as in the earlier study) was 66.5 sec. to reach a criterion of two successful repetitions of the correct response. Adding the new group to the previous five groups did not significantly change the *F* value originally reported. The present *F* of .826 is only slightly lower than the previously reported *F* of .845. The *t* between the new mean of 66.5 and the mean for the previous 100 *S*s of 61.5 is .3, well below the *t* of 1.9 needed for a significant difference at the .05 level.

It is interesting to note that *S*s needed a mean of about eight trials to complete the learning. This figure suggests that they were allowing themselves about 5 sec. per response on the average. At first *S*s appeared somewhat uncertain about the freedom allowed in the procedure and "experimented" about, trying long times and short times. It had been anticipated that *S*s might hesitate or delay the action until they were sure of a response; this, however, did not occur. They apparently were under some self-imposed instruction to repeat the material.

TABLE 1

SELF-TIMING *S*S COMPARED WITH *S*S WITH VARYING AMOUNTS OF TIME ASSIGNED TO THE RESPONSE SYLLABLE IN PAIRED-ASSOCIATE LEARNING

Condition Response Exposure Time	Mean Trials to Criterion	*SD*	Mean Total Time	*SD*
2 sec.	10.2	4.1	61.2	24.3
4 sec.	8.8	3.8	70.1	30.4
Variable[a] (*S* controlled)	8.1	2.4	65.5	26.8
6 sec.	5.8	1.9	57.9	19.0
8 sec.	4.7	2.5	56.1	29.6
15 sec.	3.3	1.2	62.2	22.2

Note.—$N = 20$ per condition.
[a] Present study. The other data are from the earlier report.

---

[1] This study was supported by United States Public Health Research and Fellowship Branch Grant M-3489.

The results of the present study, based on a self-timing procedure, support the earlier findings with *E*-controlled timing. It appears that in learning paired-associate nonsense syllables *S*s take a fixed (within limits) amount of time to learn a fixed amount of material even though they control presentation time, and thus, indirectly, the number of trials.

REFERENCE

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *J. exp Psychol.*, 1962, **63**, 409–412.

(Received June 25, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 617–618

PERFORMANCE IN EYELID CONDITIONING FOLLOWING CHANGES IN REINFORCEMENT SCHEDULE[1]

WILLARD N. RUNQUIST[2]
Pomona College

4 groups of 16 Ss were run in a classical eyelid conditioning experiment. 2 groups received 76 trials under a random 50% reinforcement schedule, 1 group with a 50-mm. airpuff as the UCS and the other group with a 100-mm. airpuff as the UCS. The other 2 groups received 60 trials of 100% reinforcement followed by 16 trials of 50% reinforcement. The same 2 UCS intensities were used in these 2 groups. Acquisition data showed little effect of UCS intensity although partial reinforcement resulted in lower performance. Following the shift, performance dropped in the continuously reinforced groups, and there was little evidence of any overall difference in postshift performance. A further analysis of the trials following nonreinforcement suggested that nonreinforcement produced more inhibition when the UCS was stronger.

Ross (1959) reports that changes in performance in classical eyelid conditioning are almost immediate when reinforcement schedule is shifted from continuous to partial reinforcement. This experiment extends the findings of Ross by determining the effect of this shift in reinforcement schedule under slightly altered conditions; namely, no ready signal, omission rather than delay of the UCS on nonreinforced trials, longer CS-UCS interval and UCS duration, and two separate UCS intensities.

Method.—The conditioning laboratory was physically similar to the Iowa laboratory. The system for presenting stimuli and recording eyeblinks was basically the same as that described by Spence and Taylor (1951) except that a Brush Mark II recorder was used. In addition, all trials were automatically programed by means of a cam driven by a 2-rph motor. All *S*s were run in a semidarkened room in which an electric fan provided masking noise. The CS consisted of the lighting of a 42-mm. diameter, milk glass covered aperture by a Drake 105 neon pilot lamp. The UCS was an air puff delivered to the right eye 560 msec. following CS onset and 400 msec. in duration.

The experiment utilized a 2 × 2 factorial design with UCS intensity and reinforcement schedule being the two orthogonal variables. Six male and 10 female college students comprised each group. The two UCS intensities were 50 and 100 mm. of mercury. The two reinforcement schedules were a random 50% reinforcement and a 100% reinforcement changed to a random 50% reinforcement. The four groups thus formed were designated weak-partial (W-P), strong-partial (S-P), weak-change (W-CH), and strong-change (S-CH).

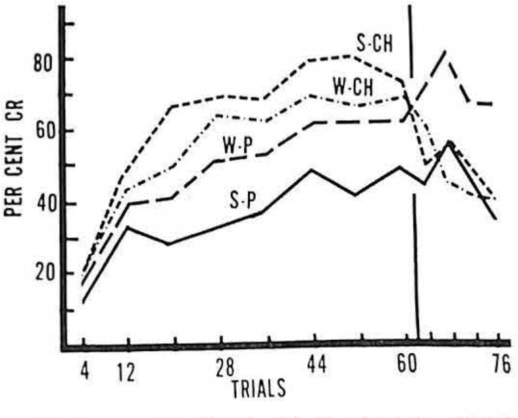


FIG. 1. Percent CRs for blocks of trials. (Trials 1–4 and 61–76 are in blocks of four trials. The other points are blocks of eight trials. The vertical line indicates the switch to partial reinforcement.)

---

[1] Partially supported by a grant from the Claremont Graduate School.

[2] Now at the University of Alberta.

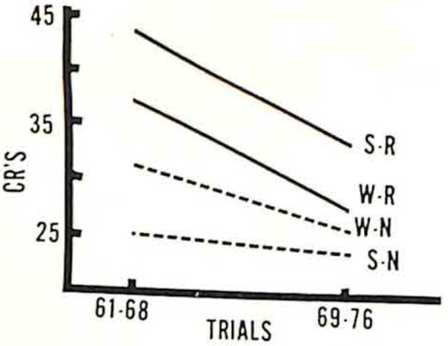


FIG. 2. Total CRs following reinforced and nonreinforced trials for the CH groups in Trials 61–76.

The conditioning procedure consisted of 76 trials, randomly spaced, with a mean intertrial interval of 22 sec. and a range of 15–32 sec. No ready signal was used. In the P conditions, the UCS was omitted on a random 50% of the trials. In the CH conditions, 60 paired CS-UCS trials were followed by 16 trials in which the UCS was omitted on a random 50%. The CRs were recorded from 200 to 525 msec.

Results.—Curves showing the percent CRs by blocks of trials are presented in Fig. 1. The results are generally consistent with the findings of Ross (1959) concerning changes in reinforcement schedule and with those of Ross and Spence (1960) concerning the effects of UCS intensity with partial reinforcement schedules. The somewhat large difference favoring the W-P group over the S-P group is discrepant from the Spence and Ross findings, but probably can be discounted as having theoretical importance. Evidence was obtained that this group contained a disproportionate number of *S*s with high scores on the Taylor Manifest Anxiety scale. Since this variable has been shown to be related to conditioning performance, it is likely that the difference obtained here represents sampling error with respect to anxiety score and, hence, conditionability.

Analysis of variance of the preswitch data, using sex as a third component of the analysis, produced significance only for reinforcement schedule ($F = 5.43$), with the apparent interaction of schedule with UCS intensity failing to be significant ($F = 1.87$). All other *F*s were less than 1.00. (All comparisons were for 1 and 56 *df*, with an error *MS* of 348.45.) Analysis of variance of the postswitch data resulted in no significant differences. The *F*s greater than 1.00 were for UCS intensity (1.74), reinforcement schedule (2.13), and the interaction of the two (2.13; all comparisons with 1 and 56 *df*; error *MS* of 29.12). Assessing the rate of performance drop is extremely difficult due to the extreme variability in the data; nevertheless, the drop appears to be more rapid in the S-CH group (27% in the S-CH vs. 5% drop in the C-CH group from Trial Block 57–60 to 61–64). In the former group, 10 *S*s showed a loss and none a gain following the introduction of nonreinforced trials, while 6 W-CH *S*s showed a loss and 4 a gain. Thus, the drop appears to be more rapid with the stronger UCS.

Further assessment of the rate of change of performance following the introduction of nonreinforced trials was made by comparing the number of CRs occurring on trials following the occurrence of a UCS with the trials following nonoccurrence of a UCS. These results are shown in Fig. 2 for the CH groups. Performance is clearly higher following reinforced trials for both W and S groups, although the difference is greater in the S group. Essentially, the same findings have been reported by Spence and Trapold (1961) for a single UCS intensity. Although statistical analysis of this trend appears to demand an overall analysis, this was precluded by extreme skewness and a restricted range of values. Instead, some attempt was made to verify the differential effect of UCS intensity by utilizing a simple chi square test on the difference scores (CRs following reinforced minus CRs following nonreinforced trials). Comparing the proportion of positive differences in the S-CH and W-CH groups resulted in a chi square of 4.50 which was significant beyond the 5% level. A final analysis is designed to show that the effect of UCS intensity depends upon previous reinforcement schedule. The same analysis of difference scores was made on the first 16 and the last 16 trials in the partial reinforcement groups. Although there was some suggestion of greater difference with the stronger UCS, the obtained chi square values were both far from significant (.75 and .39). There is evidence, then, that the inhibition generated by a nonreinforced trial is greater following omission of a strong UCS than following omission of a weak UCS, and that this effect is dependent to some extent on the past reinforcement schedule.

REFERENCES

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57,** 74–82.

ROSS, L. E., & SPENCE, K. W. Eyelid conditioning performance under partial reinforcement as a function of UCS intensity. *J. exp. Psychol.*, 1960, **59,** 379–382.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.

SPENCE, K. W., & TRAPOLD, M. A. Performance in eyelid conditioning as a function of reinforcement schedules and changes in them. *Proc. Nat. Acad. Sci., Washington, D. C.*, 1961, **47,** 1860–1868.

(Received July 2, 1962)

Journal of Experimental Psychology
1963, Vol. 65, No. 6, 619–620

INFLUENCE OF THE WRITTEN RECALL MEASURE ON FIRST-LIST ASSOCIATIONS

JUDITH GOGGIN
University of California

An experiment using A-B, A-C and A-B, A-B′ paradigms was done to ascertain if first-list responses are available even though the associations are extinguished. The recall task was designed to maximize response recall. Response recall for both lists was higher in the A-B, A-B′ group. Considering the stringent scoring method alone, differences between lists, paradigms, and interaction between lists and paradigms were significant. The general results confirm Barnes and Underwood's (1959) conclusion that both the associations are extinguished and the responses are unavailable after learning the A-C list.

In the study of Barnes and Underwood (1959) comparing transfer and retroactive inhibition under the A-B, A-C and A-B, A-B′ paradigms, the results suggested that in the A-B, A-C case, not only are the first-list associations extinguished but also the responses (B) per se are not available. The investigators pointed out, however, that this conclusion could not be reached with complete confidence because of the nature of the instructions to *S*s and the form of the recall task. The *S*s were required to reproduce both the first-list response (B) and the second-list response (C) to successive stimulus items (A). The present study is concerned with the availability of responses when the instructions and written recall task are designed to maximize *S*s' recall of these responses, independently of the retention of the specific associations.

Method.—Two groups of 16 *S*s each were used. The two successive lists learned by one group conformed to the A-B, A-B′ paradigm and those learned by the second group to the A-B, A-C paradigm. The stimulus items were nonsense syllables of from 47% to 53% association value (Glaze, 1928). Low intralist similarity was achieved by using four vowels twice and each consonant only once. The responses in the A-B, A-B′ paradigm, identical to those used by Barnes and Underwood (1959), were two-syllable adjectives taken from Haagen (1949). High interlist response similarity was obtained by pairing synonymous adjectives with the same stimulus item. There was low intralist response similarity. In the A-B, A-C paradigm, the first-list responses (B) were the same as those in the A-B, A-B′ paradigm. The second-list responses (C) were matched with the B′ response items insofar as they had similar frequencies of occurrence (Thorndike & Lorge, 1944) and the same initial letters. There was low intralist response similarity and interlist response similarity was low in that none of these adjectives came from the same clusters as the first-list responses (Haagen, 1949). The lists were presented at a 2.2-sec. rate on a memory drum with an intertrial interval of 8 sec. The *S*s learned the first list to one perfect recitation. They were then given the second list for 20 anticipation trials. Two different pairings of stimuli and responses were used and an equal number of *S*s in each group was given each pairing. The presentation of the two lists was separated by a 2-min. interval. After 20 trials on List 2, *S*s were given a blank piece of paper on which to write all the responses they could recall. The list of stimuli was typed in alphabetical order on another sheet of paper, and this sheet was available to *S*s throughout the test. The *S*s were instructed to write down all the responses they could remember in whatever order they came to mind. They were told that the list of stimuli could be used, if need be, to help in recalling the responses, but they were urged not to go down the list, syllable by syllable, and match the responses from the two lists. The *S*s were allowed as much time as they needed for this task. At the end of the test they were instructed to write beside the response the stimulus with which it was associated and to indicate whether the response came from List 1 or 2. The *S*s were told that they could add any other responses remembered at this time.

Results.—The mean number of trials to one perfect recitation of List 1 was 13.06 for both groups combined. The mean for the A-B, A-B′ group was 11.00 and that for the A-B, A-C group was 15.13. After a log transformation, $t = .34$. The total correct responses for the first three trials on Lists 1 and

2 were compared for both groups as a measure of net transfer effects. For the A-B, A-B′ group, the mean number correct on List 1 was 6.44 and on List 2 was 13.81; for the A-B, A-C paradigm, the two means were 4.69 for List 1 and 5.25 for List 2. An analysis of variance showed that the differences between paradigms and lists and the interaction between paradigms and lists were significant, $Fs(1, 30) = 28.68$, 28.03, and 20.65, respectively ($p < .01$). The number of overt intrusions was greater for the A-B, A-B′ group than for the A-B, A-C group. Table 1 shows the number of correct responses recalled by each group, using the same two scoring methods as Barnes and Underwood (1959). In the first method of scoring, a response was considered correct if it was paired with the right stimulus item and given the correct list number. In Method 2, a response was counted correct if it was recalled, whether or not the stimulus with which it was paired and the list number were remembered. For both methods of scoring, recall of List 2 responses was slightly higher for the A-B, A-B′ group; recall of List 1 responses was greater for the A-B, A-B′ group than for the A-B, A-C group. The differences between lists, paradigms, and the interaction between lists and paradigms were all significant when the stringent method of scoring alone is considered, $Fs(1, 30) = 58.18$, 21.00, and 14.54, respectively ($p < .01$). Difference scores between the two scoring methods were calculated after a Freeman-Tukey transformation. Here the difference between paradigms is significant, $F(1, 30) = 4.47$ ($p < .05$). This appears to indicate that, in the case of a few *S*s, even though an association is extinguished, the response may be available for recall. The general results lead to the conclusion that Barnes and Underwood (1959) were justified in suggesting that, in general, not only are the A-B associations extinguished during the learning of the A-C list, but also that the responses themselves are not available. In the present experiment, *S*s were urged to put down all the responses they could remember without any specification that they were to match them with the stimuli. Thus, if an *S* could remember a response without recalling the stimulus with which it was paired, he had an opportunity to put it down. In spite of this, *S*s did not recall any more responses than they did when the method of written recall was such as to implicitly stress the pairing of stimuli and responses.

TABLE 1

Mean Number of Correct Responses Scored by Two Methods

Method	List	A-B, A-B′ Group	A-B, A-C Group
1	1	6.75	4.19
	2	7.75	7.18
2	1	6.94	4.81
	2	7.94	7.69

References

Barnes, J. M., & Underwood, B. J. "Fate" of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58**, 97–105.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Haagen, C. H. Synonymity, vividness, familiarity, and association-value ratings for 400 pairs of common adjectives. *J. Psychol.*, 1949, **30**, 185–200.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received August 8, 1962)

ERRATUM

In the article "Effect of Continued Nonreinforcement on the Frustration Effect" by James H. McHose (*J. exp. Psychol.*, 1963, **65**, 444–450), the captions for Fig. 1 and 2 on pages 446 and 449, respectively, are correct and follow text, but the figures have been interchanged.